Studying the Differential Co-expression of MicroRNAs Reveals the Significant Role of White Matter in Early Alzheimer's Progression (Supplementary details)

Malay Bhattacharyya and Sanghamitra Bandyopadhyay*
Machine Intelligence Unit, Indian Statistical Institute,
203 B. T. Road, Kolkata - 700108, India
E-mail:{malay_r, sanghami}@isical.ac.in.

November 26, 2012

Introduction

This supplementary file describes many methodological and additional experimental details that encompasses a significant portion of back-end analyses for the study on region-specific association of microRNAs (miRNAs) and Alzheimer's disease (AD). By the term 'region-specific', we indicate specific analysis in gray matter (GM) and/or white matter (WM) of the human brain.

1 Alzheimer's Disease (AD)

For the completeness of the study, we have briefed the basic details about the generation of AD, their causal biogenesis and how miRNAs might be involved in their pathways. This section is a reflection of what we know as the basic causes of AD progression, which is a complex and prominent neurogenerative disorder.

The AD is characterized by the progressive loss of memory and several other characterized functions. It is the most common form of age related cognitive impairment [4, 10]. AD is caused by the accumulation of extracellular plaques of amyloid-beta (Aβ) peptides and intracellular neurofibrillary tangles composed of hyperphosphorylated microtubular protein tau [8]. Aβ is a naturally occurring, predominantly 40 amino acid long polypeptide (Aβ-40) derived from the larger amyloid precursor protein (APP), which is an integral membrane protein expressed in many tissues and concentrated in the synapses of neurons. Though the function of APP is not known clearly, it has been found to take part in the regulation of synapse formation, neural plasticity and iron export.

A neuron, the basic unit of a nerve cell, is subdivided into the two parts – axon and dendron. The axon part is composed of microtubule bundles. Tau protein, bound to these bundles help the microtubules in the transfer of synaptic information from one neuron to another through the synapse. In diseased condition, these tau proteins attain abnormal form, lose function and detach from the bundles causing microtubule subunits to fall apart [1, 9, 22]. Released tau proteins get clumped together to form tangles. The formation of plaques (Fig. 1) involves sequential cleavages by two enzymes. First, an enzyme, β-secretase (BACE) cleaves APP to form soluble extracellular fragment. This is followed by the γ-secretase, cleaving within the transmembrane domain. This releases Aβ, which on accumulation form plaques. In normal brains, α-secretase, another enzyme, cleaves closer to the cell membrane and prevents eventual generation of Aβ peptide. Increase in the proportion of the longer, more neurotoxic form, Aβ-42, results in the formation of higher order aggregates and subsequently, plaque deposition [6, 7].

*To whom correspondence should be addressed.

Several experiments showed that treatment of primary neurons with Aβ-42 evokes a strong change in miRNA profiles with a substantial portion of miRNAs being downregulated and thus proving the fact that neuronal miRNA expression changes upon exposure to Aβ [18, 13]. The peptide Aβ initially causes a downregulation of miRNAs in hippocampal neurons. Notably, hippocampus is a region of brain where memories are stored.

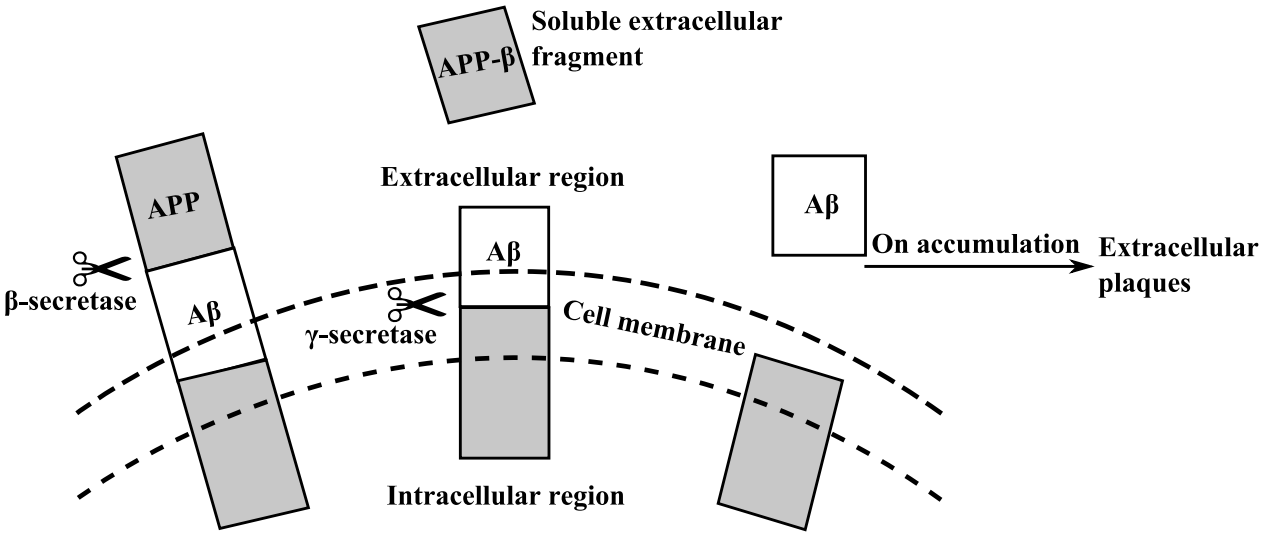

Figure S1. Formation of plaques from the accumulation of Aβ-42 in the extracellular region of the brain affected by AD.

2 MiRNAs Expressed in Brain

The activity of miRNAs is known to be tissue-specific [11]. Therefore, it is necessary to consider tissue-specificity of an miRNA while studying its functional activity. As we are interested in the activity of miRNAs in AD brains, we pursue literature mining through the PubMed [15] to accumulate the list of miRNAs expressed in the human brain tissue. For this, first we survey the literature to make a list of brain-specific miRNAs and additionally explored the miRNAs known to have association with brain-specific diseases like AD, dementia, Schizophrenia, etc. We finally obtain more than 125 miRNAs expressed in brain that is quite a large fraction of the total 1527 miRNAs listed in the latest repository (release 18) of miRBase [16].

3 Reference to an In Press Article

Very recently, Schonrock and his colleagues have carried out an extensive literature survey to enquire about the complex network of miRNAs surrounding the regulation of the amyloid 35 precursor protein (APP), which is a causal factor for AD [13]. They report about more than 30 miRNAs (listed in the last column of Table S1) taking part in the direct progression of AD and classify their activity into the categories – de-regulation in AD brain, direct regulation of the APP mRNA, regulation of APP alternative splicing, indirect regulation of APP processing, down-stream effects of Aβ on miRNA expression, with some conclusive roles in specific.

4 Gold Standard List of AD-associated MiRNAs

We have carried out extensive literature survey to accumulate the list of miRNAs having known association with AD based on biological validations. Most of these miRNAS are taken from the Human MiRNAs &

AD related miRNAs [12]	AD related miRNAs [21]	miRNAs with aberrant expression in AD brains [17]				AD related miRNAs [13]	
miR-107	miR-124	miR-9	miR-21	miR-100	miR-425	miR-124	miR-298
miR-29a	miR-125b	miR-128	miR-222	miR-212	miR-30e-5p	miR-106a	miR-328
miR-29b-1	miR-103	miR-146a	miR-9-1	miR-363	miR-92	miR-520c	miR-103
miR-106b	miR-107	miR-146b	miR-9-2	miR-125b	miR-200c	miR-20a	miR-107
miR-146a	miR-15a	miR-29a/b-1	miR-92b	miR-511	miR-423	miR-106b	miR-181c
miR-17	miR-15b	miR-15a	miR-9-3	miR-320	miR-30c	miR-17-5p	miR-148b
miR-20a	miR-16	miR-27a	miR-34a	miR-27b	miR-18b	miR-101	miR-30c
miR-21	miR-195	miR-19b	miR-326	miR-34a	miR-615	miR-147	miR-20b
	miR-424	let-7i	miR-129-1	miR-145	miR-629	miR-655	miR-361
	miR-128	miR-101	miR-129-2	miR-148a	miR-637	miR-323-3p	miR-21
	miR-29a	miR-106b	miR-136	miR-381	miR-657	miR-153	miR-409-3p
	miR-29b	miR-22	miR-181c	miR-422a		miR-29a	let-7i
	miR-29c	miR-26a	miR-197	miR-98		miR-29b	miR-125b
	miR-214	miR-26b	miR-210	miR-132		miR-9	miR-128a
						miR-29c	miR-15

Table S1 The list of miRNAs related to AD as obtained by literature mining.

Diseases Database (HMDD). However, we have additionally done literature mining based on PubMed [15] to add several other miRNAs that have altered or aberrant expression while in association with AD [17, 21, 13], which contributes to a list of 74 distinct miRNAs. We list up the AD-associated miRNAs in Table S1. Interestingly, we found maximum support from the literature reporting the evidences of the association of miR-29a/b with AD.

5 Dataset Details

The cerebral cortex is known to contain two major components or layers. One of them is the GM which is 4mm thick and overlies above the other layer called WM containing myelinated axons. In the current study, we work on a recently published dataset which reports miRNA expressions in GM and WM in the temporal cortex of normal and early Alzheimer's-affected human [21]. The dataset contains the expression profiles of 170 miRNAs for 10 individuals (brain samples of elderly females). Neuropathological lesions such as neuritic amyloid plaques (NPs) and neurofibrillary tangles (NFTs) are found in GM predominantly. But several studies show that WM perturbations also have an effect on AD pathogenesis. For understanding the significance of both these regions in AD pathology, the gene enrichment analysis on the expression profiles in GM and WM are considered as case and control phenotypes, respectively and vice-versa.

5.1 Expression Profiling

To prepare the dataset, miRNA profiling was performed in GM and WM separately using Exiqon Locked Nucleic Acid (LNA) microarrays [21]. The expression levels of each miRNAs were obtained for all the 10 samples. Among the 10 samples, RNA quality was found to be poor for one sample and was considered degraded (Case 7D). This specific case was excluded from further analysis. The dataset provides the p-values obtained from t-tests which are used to find whether miRNAs are enriched in gray matter or white matter. It also reports correlation coefficients of the expression of miRNAs in gray matter and white matter with the counted AD lesions (DPs, NPs and NFTs). Total 50 of the 170 miRNAs we studied are present in the Gold standard list (vide Table 1). Among the 170 miRNAs in total, hsa-miR-144 and hsa-miR-422a are found to possess the highest expression values in GM and WM, respectively.

5.2 Study Group Details

Wang *et al.* were motivated to study the continuum of disease from "no AD" to "early AD" pathology. They worked on 10 subjects. Statistically, may be this number is small, but biologically the subjects are significant enough. The choice of subjects greatly depends on some physiological constraints. They were chosen such that no evidence of frontotemporal dementia, indication of cancer in the brain parenchyma, argyrophilic grains, cortical Lewy bodies, large infarctions in the brain, or microinfarcts are found within 3 cm of the brain tissue samples [21]. The postmortem interval was considered as less than 4 hours. All ten

Gray matter	White matter	ADC between GM and WM	miRNA pairs
-0.93	0.89917	1.8292	miR-423-5p & miR-92b
-0.88925	0.92444	1.8137	miR-423-5p & miR-320c
-0.93291	0.85747	1.7904	miR-423-5p & miR-140-3p
-0.8598	0.92804	1.7878	miR-423-5p & miR-320b
-0.88943	0.89664	1.7861	miR-423-5p & miR-93
-0.8633	0.92101	1.7843	miR-423-5p & miR-320a
-0.84207	0.92723	1.7693	miR-423-5p & miR-320d
-0.85702	0.8588	1.7158	miR-423-5p & miR-106b
-0.90589	0.79592	1.7018	miR-423-5p & let-7a
-0.84262	0.85577	1.6984	miR-423-5p & miR-374b
-0.83412	0.85322	1.6873	miR-423-5p & miR-15a
-0.89411	0.78247	1.6766	miR-423-5p & miR-101
-0.77331	0.8993	1.6726	miR-423-5p & miR-19b
-0.76717	0.9007	1.6679	miR-423-5p & miR-19a
-0.85493	0.8041	1.659	miR-423-5p & miR-186

(a)

miRNA	Degree	AD related
miR-423-5p	15	√
miR-92b	1	
miR-320c	1	√
miR-140-3p	1	
miR-320b	1	√
miR-93	1	√
miR-320a	1	√
miR-320d	1	√
miR-106b	1	√
let-7a	1	
miR-374b	1	
miR-15a	1	√
miR-101	1	√
miR-19b	1	√
miR-19a	1	
miR-186	1	

(b)

Table S2 (a) The top 15 miRNA pairs found to be differentially co-expressed through correlation-based analysis. (b) The degree of connectivity found by unifying the top 15 pairs, resulting in 16 miRNAs, of differentially co-expressed miRNAs.

individuals were elderly (ages between 84-92, median = 87 years) Caucasian females. However, one of the samples was excluded due to low RNA integrity numbers that indicates substantial RNA degradation. Final examinations were administered for a spectrum of 10.9 months on an average (median = 7.2 months) before death.

6 Correlation-based Analysis

Correlation is a measure of the relation between two or more variables. A coefficient of correlation is a mathematical measure of how much one variable (here, miRNA) can be expected to be influenced by changes in another. Variables can be positively or negatively correlated, which is reflected by the sign of the correlation coefficients. There are several correlation coefficients, often denoted as ρ or r, measuring the degree of correlation. The most common of these is the Pearson's correlation coefficient, which is sensitive only to a linear relationship between two variables.

Let us assume that we have a series of n measurements of the variables x and y represented as x_i and y_i, $i = 1, 2, ..., n$. Then, the standard Pearson correlation coefficient ρ between x and y can be calculated as

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{[\sum_{i=1}^{n}(x_i - \overline{x})^2][\sum_{i=1}^{n}(y_i - \overline{y})^2]}}. \quad (1)$$

Here, $\overline{x}$ and $\overline{y}$ denote the mean values over all the measurements of x and y, respectively. A Pearson correlation coefficient value of +1 indicates a positive correlation and a -1 indicates a negative correlation (anti-correlation) between the variables. The value ranges within [-1, +1] indicating the degree of linear dependence. If the correlation coefficient value approaches to zero, it means to get closer to 'uncorrelated' and if it is closer to -1 or +1, it means there is strong correlation between the variables.

We compute the Pearson correlation coefficient value separately in GM and WM for all possible pairs of miRNAs. Then we obtain the absolute difference of these two region-specific correlation values for denoting the degree of differential co-expression. Based on this value, we calculate the top 15 miRNA pairs having highest differential co-expression (shown in Table S2). It is intuitive that if any of the miRNAs in a pair shows aberrant expression then the co-expression pattern will change. So, most possibly any of those miRNAs in a pair shows a change of pattern for those miRNA pairs obtained.

We note that there is a common presence of miR-423-5p in all these miRNA pairs indicating it has some aberrant expression that is causing differential co-expression while analyzing it in pair with other miRNAs. We further exclude this from the dataset and observe that all these high connectivity of miRNAs gets disrupted. See the revised top 15 list after excluding miR-423-5p in Table S4. So, the significance of this miRNA (miR-423-5p possibly working as a hub) emerges from these cross-verifications. It is interesting

Gray matter	White matter	ADC between GM and WM	miRNA pairs
0.74853	-0.86979	1.6183	let-7b & miR-124
-0.74116	0.87369	1.6149	miR-584 & miR-32
0.75757	-0.85478	1.6124	miR-151-3p & miR-124
0.74786	-0.84152	1.5894	let-7c & miR-124
-0.76673	0.81482	1.5815	miR-584 & miR-320a
0.86792	-0.70571	1.5736	let-7c & miR-129-5p
-0.65967	0.88952	1.5492	miR-584 & miR-19b
-0.67765	0.86914	1.5468	miR-584 & miR-93
0.82359	-0.71607	1.5397	mIR-340 & miR-124
-0.74111	0.79843	1.5395	miR-584 & miR-140-3p
0.68084	-0.85549	1.5363	miR-181b & miR-124
-0.61584	0.91985	1.5357	miR-584 & miR-19a
-0.72895	0.80356	1.5325	miR-185 & miR-19a
-0.80215	0.72625	1.5284	miR-34c-5p & miR-320a
-0.67739	0.84804	1.5254	miR-584 & miR-320b

(a)

miRNA	Degree	AD related
miR-584	7	
miR-124	5	√
let-7c	2	
miR-320a	2	√
miR-19a	2	
let-7b	1	
miR-32	1	
miR-151-3p	1	
miR-129-5p	1	
miR-19b	1	√
miR-93	1	
miR-340	1	
miR-140-3p	1	
miR-181b	1	
miR-185	1	
miR-34c-5p	1	
miR-320b	1	√

(b)

Table S3 (a) The top 15 differentially co-expressed miRNA pairs found through correlation-based analysis by excluding miR-423-5p from the dataset. (b) The degree of connectivity found by unifying the top 15 pairs, resulting in 17 miRNAs, of differentially co-expressed miRNAs.

miRNA	Highest ADC value	Correlated miRNA	AD related miRNAs
miR-21	1.4481	miR-423-5p	√
miR-222	1.1796	miR-15a	√
miR-9-1	1.329	miR-19a	
miR-9-2	1.329	miR-19a	
miR-9-3	1.329	miR-19a	
miR-107	1.0347	miR-19b	√
miR-29a	1.3142	miR-15a	√
miR-29b-1	1.0448	miR-15a	√
miR-103	0.87759	miR-19b	√
miR-15b	1.5464	miR-423-5p	√
miR-195	1.1788	let-7c	
miR-29c	0.9406	let-7c	

Table S4 The 15 miRNA pairs having the highest ADC values.

to note that several miRNA pairs in Table S3(a) having high ADC values are due to positive correlation in GM and negative correlation in WM, a pattern which is opposite to the one observed from Table S5(a). The degree values of connectivity of the newly selected miRNAs in this approach are shown in Table S3(b). Among the 15 miRNA pairs found in the analysis excluding miR-423-5p, the GM-enriched miRNA miR-124 is seen to be a promising (with a high degree of connectivity with the others) and already validated candidate for AD. However, the drop in the significance level (as realized from the computed p-values) of findings based on the Gold standard from the analyses with and without miR-423-5p clearly rejects our initial doubt.

We further evaluate the behavior of miRNAs, present in the Gold standard list but not identified by our correlation-based analysis. The intuition is to check how these known AD-associated miRNAs correlates with the other miRNAs. In Table S4, we list up these miRNAs and the miRNAs with which it pairs up to reflect highest absolute difference in correlation (ADC) in GM and WM. These miRNAs have not been selected in our studies due to the stringent constraint of considering only the top 15 miRNAs in every analysis. The miRNAs otherwise neglected are in fact have close ranks in the analyses carried out and also pair up with those related to AD (see Table S4).

Instead of focusing only on the ADC values, we further pursue degree-based analysis on the entire set of miRNAs. From the 170 miRNAs, the top 15 miRNAs having the highest degree values are found by considering only those miRNA pairs having a value of ADC > 1. The degree value indicates that with how many miRNAs a particular miRNA pairs to give an ADC value greater than 1. We consider an interaction as valid if it passes the threshold ADC > 1. The total interaction count of an miRNA with the others represents its degree value. In Table S5, we show the degree values of all the miRNAs analyzed from the dataset. Total 8 among the top 15 miRNAs are found to be matching (p-value = 3.74E-02) with the Gold

miRNA	Degree	AD related
miR-423-5p	87	√
miR-84	71	
miR-94c-5p	65	
miR-129-5p	52	
miR-124	51	√
miR-19b	49	√
miR-15a	46	√
miR-299-5p	45	
miR-379	45	
miR-320b	43	√
let-7c	43	
miR-218	41	
miR-19a	40	
miR-320c	39	√
miR-129-3p	39	

Table S5 The miRNAs having highest degree values based on a filtration over the ADC values.

Table S6 An miRNA-miRNA association matrix

	miR-X1	miR-X2	...	miR-Xn
miR-X1	-	(GM-,WM+)	...	(GM+,WM-)
miR-X2	(GM-,WM+)	-	...	(GM+,WM-)
...	...	...	...	...
miR-Xn	(GM+,WM-)	(GM+,WM-)	...	-

standard list of AD related miRNAs. The p-value of this finding is not very low. Therefore, it becomes clear that finding out more significant differentially expressed miRNAs based on only higher connectivity values (degree) may not be proper. So, the analysis should consider the degree of differential co-expression and also the degree of connectivity.

7 Statistical Significance Analysis of the DCSTs

To verify the significance of the substructures (in the form of DCSTs) obtained by applying the proposed methodology, it is worth exploring the probability of receiving those by chance. So, the computation of p-value of obtaining the DCSTs is of importance here and the principle criteria is the sizes of those DCSTs. So, the question is can we obtain such large DCSTs with probabilistic chances. Randomization models that preserve the degree distribution in a network (by edge-swapping) has been effectively used in earlier studies [2, 19]. Here, we construct an miRNA-miRNA association matrix, of size $n \times n$, where n denotes the number of miRNAs retained from the primary differential co-expression network by setting $T = 1$. An element in the position (i, j) is "1" if miRNA i is differentially co-expressed with miRNA j (i.e., the ADC value between them is more than 1), otherwise "0". This basically represents a network and one such example is shown in Table S6.

Now, the edges of the matrix are perturbed repeatedly to make it random. In this randomization subroutine, the values are actually swapped randomly between two miRNA pairs to keep the distribution of the values same in the matrix. After preparing the final random matrix, the *switching patterns* are randomly perturbed. Finally, we find the DCSTs of two different patterns from this randomized matrix (network). Based on this, we generate a p-value that encounters the frequency of exceeding the sizes of the original DCSTs obtained by using the proposed method. A poor p-value computed from a solution establishes the truth of not receiving the result by chance. We have done this randomization 1000 times and obtained a very low p-value. The result establishes that the obtained DCSTs are worth of further analysis because they encompass information that is more than any random case.

Gray Matter			White Matter		
p-value	miRNA	AD related	p-value	miRNA	AD related
8.75E-05	miR-129-5p	√	6.28E-06	miR-20a	√
2.15E-03	miR-485-3p		1.06E-05	miR-219-2-3p	
5.28E-04	miR-34a	√	1.34E-05	miR-338-3p	
5.37E-04	miR-143		1.17E-05	miR-17	√
7.32E-04	miR-149		1.63E-05	miR-106a	
1.00E-03	miR-136	√	1.78E-05	miR-16	√
1.07E-03	miR-138		2.10E-05	miR-219-5p	
1.38E-03	miR-145	√	2.43E-05	miR-32	
1.40E-03	miR-124	√	3.43E-05	miR-19a	
1.61E-03	miR-378		4.46E-05	miR-584	
1.62E-03	miR-129-3p	√	5.07E-05	miR-338-5p	
1.62E-03	miR-432		7.33E-05	miR-151-5p	
1.67E-03	miR-29b	√	7.89E-05	miR-23b	
1.70E-03	miR-488		9.50E-05	miR-34c-5p	
1.84E-03	miR-128	√	1.11E-04	miR-27b	√

Table S7 The AD-association of differentially expressed miRNAs enriched in GM and WM as obtained by applying the paired student's t-test.

8 Paired Student's t-test

We apply the paired student's t-test on the expression data profiled in [21] for differential expression analysis. The paired student's t-test is performed by using the implementation in R. This test is used to compare two sets of quantitative data when data in each sample set are dependent on each other. The paired t-test is generally used when observations are taken from the same subject for two different cases (say, GM and WM here). We compute this by calculating the test statistic as follows

$$T_i = \frac{\mu_{i1} - \mu_{i0}}{\sqrt{\frac{\sigma_{i1}^2}{n_1} + \frac{\sigma_{i0}^2}{n_0}}}. \tag{2}$$

Genewise study is conducted here to compute the test statistic. For a gene i, μ_{i1} and μ_{i0} are the mean values and σ_{i1} and σ_{i0} are the standard deviations of the expression vectors in the case and control phenotype, respectively. For the current analysis, $n_0 = n_1 = 9$, and so the formula of computing the test statistic becomes

$$T_i = \frac{3(\mu_{iGM} - \mu_{iWM})}{\sqrt{\sigma_{iGM}^2 + \sigma_{iWM}^2}} \quad \text{and} \quad T_i' = \frac{3(\mu_{iWM} - \mu_{iGM})}{\sqrt{\sigma_{iWM}^2 + \sigma_{iGM}^2}}. \tag{3}$$

Here a single miRNA is considered at a time for differential expression analysis, where the mean values (μ_{iGM} and μ_{iWM}) and standard deviations (σ_{iGM} and σ_{iWM}) are for miRNA i present in GM and WM, respectively. After getting the test statistic, the p-values are computed for deriving the statistical significance of the observation. From a mathematical point of view, p-value is the probability of getting a test statistic at least as extreme as the one that is actually observed, assuming that the null hypothesis is true. In other words, it gives the probability of the getting the result by chance. If a smaller p-value is obtained, then the result is statistically significant, i.e., unlikely to have occurred by chance and the null hypothesis can be rejected. Here, the null hypothesis is that miRNAs are not explicitly enriched in GM or in WM. The computed p-values based on the test statistic for the top 15 miRNAs are shown in Table S7.

Note that, a minimum sample size around twenty can be estimated, assuming the effect size = 0.8 (Cohen's d), statistical power level = 0.8 and significance level = 0.05, for conducting t-test. But here, the sample size is restricted to no more than ten. This might be an important reason for the inefficacy of t-test for the current analysis.

9 Significance Analysis of Microarrays

Significance analysis of microarrays (SAM) is a technique for determining the changes in gene expression pattern that are statistically significant. A gene-specific test statistic is computed by SAM for evaluating the relative difference in gene expression based on permutation analysis of expression profiles [20]. The test statistic is calculated as

Gray matter miRNAs	AD related	White matter miRNAs	AD related
miR-124	√	miR-219-2-3p	
miR-129-5p	√	miR-19a	
miR-381	√	miR-338-5p	
miR-129-3p	√	miR-584	
miR-138		miR-219-5p	
miR-379		miR-338-3p	
miR-299-5p		miR-20a	√
miR-135a		miR-181b	
miR-7		miR-17	√
miR-128	√	miR-106a	√
miR-149		miR-181a	
miR-218		miR-151-3p	
miR-411		miR-32	
miR-127-5p		miR-19b	√
miR-127-3p		miR-151-5p	

Table S8 The list of differentially expressed miRNAs found by applying the SAM method.

p-value testing		Ranks obtained with SAM	
miRNA	p-value	GM (+ve genes)	WM (-ve genes)
miR-21	2.87E-04		21
miR-222	6.38E-03	26	
miR-9-1	4.55E-03	32	
miR-9-2	4.55E-03	32	
miR-9-3	4.55E-03	32	
miR-107	3.11E-03	45	
miR-29a	8.27E-03	33	
miR-29b-1	1.67E-03	37	
miR-103	1.42E-02	56	
miR-15b	1.74E-04		26
miR-195	8.07E-03	52	
miR-29c	8.30E-03	62	

Table S9 Analysis of the miRNAs known to be related to AD as per the Gold standard but not obtained with the analysis with SAM.

$$d_i = \frac{r_i}{s_i + s_0}, \tag{4}$$

where, r_i is the linear regression coefficient of gene i, s_i denotes the standard error of r and s_0 is computed as the percentile based on the level of significance (termed as alpha). The constant s_0 is chosen in a way to minimize the coefficient of variation of d_i.

We apply the SAM method, collecting the authors' version available online, to find out differentially expressed miRNAs separately in GM and in WM. The miRNAs are listed, in order of their ranks and corresponding p-values obtained with SAM, in Table S8. We can observe that the top few miRNAs differentially expressed in GM are somehow captured with SAM returning lower ranks, but in general the selection of differentially expressed miRNAs both in GM and WM are very much random (p-value > 0.1).

We also follow up the p-values and corresponding positions of the miRNAs that are listed in the Gold standard but are out of the top 15 list of significant miRNAs obtained with SAM. The results indicate that many of the miRNAs would have entered into the list of significant miRNAs if the selection threshold was higher. This is shown in Table S9. But again there may be many false positive entries with the change of threshold. As a whole we understand that differential expression analysis may not be suitable for finding groups of disease-specific miRNAs. Differential co-expression analysis [3] thus become the second option for further analysis.

10 Graph Clustering

First of all, we use the grPartition graph clustering algorithm [5], which partitions a graph based on spectral factorization. We also use the well-known generalized topological overlap measure (GTOM) [23] with grPartition for clustering the graph. Genes are said to have high topological overlap if they are connected

Cluster No.	grPartition		grPartition with GTOM		LinLogLayout	
	CC	BC	CC	BC	CC	BC
1	0.1333	0.16667	0	0.4545	0.22414	0.0883
2	0.1732	0.22727	0	0.375	0.11579	0.404
3	0	0.71428	0	0	0.21875	0.3629
4	0	0.33333	0.0074	0.2353	0.21429	0.2917
5	0	0	0.0019	0.3043	0.16667	0.3333
6	0	0.55556	0	0.0769		
7	0	0.6	0	0.1875		
8	0	0.125	0.6667	1		
9	0.2532	0.3	0	0.1818		
10	0.1667	0.25	1	0.5		

Table S10 Comparative study between different graph clustering algorithms.

to roughly the same group of genes in the network (i.e. they share the same neighbourhood). Generalizing the topological overlap measure from $m = 1$ step neighbourhoods to $m \geq 2$ step neighbourhoods, defines the m-th order GTOM. A weighted graph is obtained from GTOM and is clustered using grPartition. We use a third graph clustering algorithm, LinLogLayout, which prepares the graph layouts and produces the graph clustering results [14]. From the results obtained, the strengths of the modules have been calculated.

We compute the strength of each cluster based on two measures. The first one, clustering coefficient (CC), denotes the interaction strength of a cluster and is calculated as $CC = e/\binom{n}{2}$, where e is the number of edges present in the cluster and n is the total number of nodes. The other one, a biological coefficient (BC) for each cluster, is calculated as the ratio of the number of miRNAs related to AD (in a particular cluster) to the total number of miRNAs present in that cluster. This signifies how AD-associated an miRNA module is.

Table S10 shows the results obtained from the approaches GTOM, grPartition and LinLogLayout. Table S10 lists up the strength of the clusters in terms of the CC and BC values for each cluster. Some of these clusters having zero strength indicates that the miRNAs included in these modules do not have edges between them. As can be observed from the results, there is no dependence between the CC and BC values. Most importantly, for none of the clusters the BC value is quite good (maximum 0.4545 for GTOM). So, these graph clusters are not good candidate sets of AD-associated miRNA modules. This ineffectiveness of modularization has already been discussed while pointing out that (see Methods section in the main article) differential co-expression patterns are not transitive.

References

[1] A. C. Alonso, I. Grundke-Iqbal, and K. Iqbal. Alzheimer's disease hyperphosphorylated tau sequesters normal tau into tangles of filaments and disassembles microtubules. *Nature Medicine*, 2(7):783–787, 1996.

[2] S. Bandyopadhyay and M. Bhattacharyya. Analyzing miRNA co-expression networks to explore TF-miRNA regulation. *BMC Bioinformatics*, 10:163, 2009.

[3] S. Bandyopadhyay and M. Bhattacharyya. A biologically inspired measure for co-expression analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(4):929–942, 2011.

[4] E. M. Blalock, J. W. Geddes, K. C. Chen, N. M. Porter, and W. R. Markesbery. Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proceedings of the National Academy of Sciences USA*, 101(7):2173–2178, 2004.

[5] S. Bohacek, J. Hespanha, and K. Obraczka. Hierarchical max-flow routing. In *Proceedings of the IEEE Global Telecommunications Conference*, St. Louis, MO, 2005.

[6] M. K. Carpenter, K. A. Crutcher, and S. B. Kater. An analysis of the effects of Alzheimer's plaques on living neurons. *Neurobiology of Aging*, 14(3):207–215, 1993.

[7] D. Galimberti and E. Scarpini. Alzheimer's disease: From pathogenesis to disease-modifying approaches. *CNS & Neurological Disorders – Drug Targets*, 10(2):163–174, 2011.

[8] S. Hauptmann, U. Keil, I. Scherping, A. Bonert, A. Eckert, and W. E. Müller. Mitochondrial dysfunction in sporadic and genetic Alzheimer's disease. *Experimental Gerontology*, 41(7):668–673, 2006.

[9] K. S. Kosik, C. L. Joachim, and D. J. Selkoe. Microtubule-associated protein tau (τ) is a major antigenic component of paired helical filaments in Alzheimer disease. *Proceedings of the National Academy of Sciences USA*, 83(11):4044–4048, 1986.

[10] V. Krishnamurthy, N. S. Issac, and J. Natarajan. Computational identification of Alzheimer's disease specific transcription factors using microarray gene expression data. *Journal of Proteomics & Bioinformatics*, 2(12):505–508, 2009.

[11] M. Lagos-Quintana, R. Rauhut, A. Yalcin, J. Meyer, W. Lendeckel, and T. Tuschl. Identification of tissue-specific microRNAs from mouse. *Current Biology*, 12:735–739, 2002.

[12] M. Lu, Q. Zhang, M. Deng, J. Miao, Y. Guo, W. Gao, and Q. Cui. An analysis of human microRNA and disease associations. *PLoS ONE*, 3(10):e3420, 2008.

[13] L. M. Ittner N. Schonrock, M. Matamales and J. Götz. MicroRNA networks surrounding APP and amyloid-β metabolism - implications for alzheimer's disease. *Experimental Neurology*, in press, 2012.

[14] A. Naock. Energy models for graph clustering. *Journal of Graph Algorithms and Applications*, 11(2):453–480, 2007.

[15] PubMed. NCBI. `http://www.ncbi.nlm.nih.gov/pubmed`, 2012. [Online; accessed between December 2010-January 2012].

[16] S. van Dongen S. Griffiths-Jones, H. K. Saini and A. J. Enright. mirbase: tools for microrna genomics. *Nucleic Acids Research (Database issue)*, 36:D154–D158, 2008.

[17] J. Satoh. MicroRNAs and their therapeutic potential for human diseases: Aberrant microRNA expression in Alzheimer's disease brains. *Journal of Pharmacological Sciences*, 114(3):269–275, 2010.

[18] N. Schonrock, Y. D. Ke, D. Humphreys, M. Staufenbiel, L. M. Ittner, T. Preiss, and J. Götz. Neuronal microRNA deregulation in response to Alzheimer's disease amyloid-β. *PLoS ONE*, 5(6):e11070, 2010.

[19] R. Shalgi, D. Lieber, M. Oren, and Y. Pilpel. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Computational Biology*, 3(7):e131, 2007.

[20] V. G. Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences USA*, 98(9):5116–5121, 2001.

[21] W. X. Wang, Q. Huang, Y. Hu, A. J. Stromberg, and P. T. Nelson. Patterns of microRNA expression in normal and early Alzheimer's disease human temporal cortex: white matter versus gray matter. *Acta Neuropathology*, 121(2):193–205, 2011.

[22] J. G. Wood, S. S. Mirra, N. J. Pollock, and L. I. Binder. Neurofibrillary tangles of Alzheimer disease share antigenic determinants with the axonal microtubule-associated protein tau (τ). *Proceedings of the National Academy of Sciences USA*, 83(11):4040–4043, 1986.

[23] A. M. Yip and S. Horvath. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics*, 8:22, 2007.