Molecular BioSystems

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Cite this: *Mol. BioSyst.*, 2013, **9**, 457

Studying the differential co-expression of microRNAs reveals significant role of white matter in early Alzheimer's progression†

Malay Bhattacharyya and Sanghamitra Bandyopadhyay*

MicroRNAs (miRNAs) are a class of short non-coding RNAs, which show tissue-specific regulatory activity on genes. Expression profiling of miRNAs is an important step for understanding the pathology of Alzheimer's disease (AD), a neurodegenerative disorder originating in the brain. Recent studies highlight that miRNAs enriched in gray matter (GM) and white matter (WM) of AD brains show differential expression. However, no in-depth study has yet been conducted on analysing the differential co-expression of pairs of miRNAs over GM and WM. Two genes (or miRNAs) are said to be co-expressed if their expression profiles change similarly over a number of samples. A pair of co-expressed genes under a condition type (or phenotype) may not remain co-expressed, or get contra-expressed, under another condition. Such pairs of genes are referred to as differentially co-expressed. Such an investigation in the early stage of AD is reported in this article. A network of differentially co-expressed miRNAs in GM and WM is first built. Analysis of the differential co-expression property reveals that such a network can not have any cycle. We use the notion of switching to distinguish two distinct types of differential co-expression patterns – a pair of miRNAs that are highly co-expressed in GM but does not remain so in WM, and *vice versa*. Based on this, we find the substructures, referred to as differentially co-expressed switching tree (DCST), that throughout have similar pattern of switching. The miR-423-5p emerges as a hub of the network. We extract subtrees of these DCSTs that have similar *switching pattern* throughout. These substructures are found to be both statistically and biologically significant. A large number of miRNAs obtained from the DCSTs are found to have association with AD, most of which are enriched in WM. This computational study therefore indicates a significant role of WM in early AD progression, a hitherto less acknowledged fact.

Received 25th January 2012,
Accepted 17th December 2012

DOI: 10.1039/c2mb25434d

www.rsc.org/molecularbiosystems

1 Introduction

MicroRNAs (miRNAs) are small RNAs that play major roles in both normal and diseased conditions in different organisms, by regulating gene expression through the combination with RNA-induced silencing complex (RISC).[1] RISC uses the miRNAs as a template for recognizing the complementary mRNA, and on getting a complementary strand, it activates RNase and translationally represses or cleaves the RNA. This process is important in the regulation of genes by miRNAs. The RISC complex along with miRNAs negatively regulates gene expression by degrading mRNA, inhibiting translation or by performing both of these functions.[2] These small regulatory RNAs have potential roles in metabolism, neurodevelopment, neuroplasticity, apoptosis, and other fundamental neurobiological processes.[3] In fact, there are many molecular processes that are controlled by the changes in miRNA abundance. In this study, we are interested in Alzheimer's disease (AD), a neurodegenerative disorder (see section 1 of ESI† for a precise description on AD), in which problems crop up related to memory, thinking and behaviour.[4] It is a complex and prominent age related neurogenerative disorder characterized by progressive loss of memory and several other functions. In pathological AD cases, the expression profiles of brain-enriched miRNAs have been shown to alter from the normal cases. Remarkably, a large number of miRNAs (more than 8% of the known miRNAs, see section 2 in ESI†) are found to be expressed in the brain, more so than in any other tissue type.

Microarray data has been widely used in disease analysis by studying differential expression of genes[2,5] or miRNAs,[6–8] which

Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata - 700108, India. E-mail: malay_r@isical.ac.in, sanghami@isical.ac.in

† Electronic supplementary information (ESI) available: http://www.isical.ac.in/~bioinfo_miu/AD/Suppl.pdf. See DOI: 10.1039/c2mb25434d

show significantly varying expression patterns in multiple phenotypes (*e.g.*, diseased *vs.* non-diseased). Differential expression is a property of a single gene (or miRNA) whereas the differential co-expression is a property of a pair of genes (or a pair of miRNAs). Differential co-expression is a change in co-expression of two genes (or miRNAs) between multiple conditions. In the current analysis, we focus on differentially co-expressed miRNAs enriched in the gray matter (GM) and white matter (WM) of the brain. There is no significant study that reports about differential co-expression of miRNAs in AD brains, although it is expected to provide newer biological insights. Differential co-expression analyses of miRNAs might be useful for understanding pathological behaviors in diseases,[9] for studying evolutionary conservedness of brain[10] and also for therapeutic development in AD based upon model organism analysis.[11] We collect a list of miRNAs, after careful literature survey, that have already been experimentally validated to be associated with AD. Experiments have been conducted separately to identify the miRNAs enriched in GM and WM and also to find out the miRNA pairs having high aberrant correlation between WM and GM. Here, we introduce the notion of switching to distinguish two distinct types of differential co-expression patterns – a pair of miRNAs that are highly co-expressed in GM but does not remain so in WM, and *vice versa*. Based on this, differentially co-expressed miRNAs are selected and used by a novel algorithm to construct Differentially Co-expressed Switching Trees (DCSTs). Such a tree contains miRNAs as its nodes and an edge between a pair of nodes represents a particular *switching pattern*, which is similar throughout. Finally, we validate that the large DCSTs constructed following the proposed methodology provides a subset of miRNAs that might be strongly associated with AD.

The current paper is organized as follows. Section 2 reviews the established knowledge about the earlier association of AD (see section 2 of ESI† for a brief description on AD progression) and miRNAs. The proposed method has been discussed in section 3. Section 4 describes how the proposed method applies to fulfill the motivation of the current analysis. Finally, section 5 draws some conclusions based on the analyses and the observations obtained in the preceding sections.

2 State-of-the-art

MiRNAs, acting as effectors of neuronal development and synaptic plasticity, are abundant in the brain. They play a major role in maintenance of the neuronal phenotype.[12] They are also expressed in dendrites where they regulate spine structure and are essential mediators of synapse development. Dysregulation of miRNAs leads to synaptic collapse, thus preventing the passing of signal information between neurons, that leads to neuronal disorder. MiRNAs are also required for the development as well as to maintain the function of the brain. Aβ-induced miRNA regulation may involve alterations in cellular localization that will impact the brain function.[4]

The β-site amyloid precursor protein cleaving enzyme 1 (BACE1) is a protein coding gene significantly involved in AD. BACE1 protein is an endopeptidase that cleaves the β-amyloid precursor protein to generate neurotoxic β-amyloid peptide Aβ-42. The dysregulation of BACE1 plays a major role in pathogenesis of AD.[13] It is enriched with neurons and its expression is regulated post-transcriptionally. Several studies[3,14,15] suggest that significant changes in miRNA expression (*i.e.*, upregulation or downregulation) occur in different brain regions of an AD subject, and thus showing differential expression of miRNAs among AD and normal brains. Therefore, it is found that an alteration in the miRNA expression is a significant contributing factor for AD[16] (see ESI† section 3 for more details on this in press article).

There are several approaches carried out earlier to study the expression levels of miRNAs in the brain and its relation with AD.[3,14,15,17] For this, a miRNA microarray is prepared by extracting RNA from human brain tissue, isolated from patients with different clinical and neuropathological features. Cases are separated into four groups: nondemented without plaques, nondemented with plaques, mild cognitive impairment (MCI) with moderate AD pathology, and AD. The miR-107 is found to be one of the few miRNAs whose levels consistently changes between AD and non-demented brains. It shows significant decrease in expression levels among the four groups. It has been shown by expression profiling that miR-107 contributes to the posttranscriptional regulation of BACE1.[3] In another study, miRNA expression is assessed in human cerebral cortical GM and WM across a range of AD pathology.[15] This provides the first insight into the difference between the expression of miRNAs in GM and WM of AD-affected brains. Two primary tests are done in the study. First, miRNAs enriched in GM and WM are found using paired student's *t*-tests. By clustering the data based on the similarity in expression profiles, they discriminate miRNAs specific to regions (GM and WM). This is further used to determine which miRNAs are enriched in GM and WM. Amyloid plaques are separated into DPs (plaques without neurites) and NPs (plaques with neurites). Neurites are filamentous or pointed outgrowths from the cell body of neurons.[17,18] To correlate miRNA expression data across different cases (DPs, NPs and NFTs), GM data is queried separately from WM data and analyses are done separately for all three counted AD lesions. Simple correlation coefficients are obtained. Linear regression based tests are also done on selected subsets of miRNAs and are tested for correlation with the density of AD lesions. These miRNAs are chosen on the basis that they are previously found to have altered expression in association with AD. To test whether there are group patterns in miRNA expression in the dataset, a similarity expression matrix is obtained by comparing all the 170 miRNAs using the Partek Genomic suite. This matrix is hierarchically clustered to see the patterns of similarity of miRNA expression across different individuals including both GM and WM. This clustered similarity matrix is used to separate different miRNAs into groups (A–E) related to their expression patterns across different samples.[15]

3 Methods

Differential co-expression analyses are central to disease analysis.[9,19–23] For a system level study, differential co-expression networks can be built up corresponding to a specific disease. Note that differential networks[24] are fundamentally different

from such networks. In differential co-expression networks, an edge between two nodes (biomolecules) indicates that they are differentially co-expressed. However, there remains a basic question about the transitivity of differential co-expression patterns for a pair of biomolecules in different phenotypes. Let us consider the scenario when there are associations between three biomolecules A, B and C, as shown in Fig. 1a. Assume that the biomolecule A is differentially co-expressed with B and B is differentially co-expressed with C. Then, is C differentially co-expressed with A?

Let us illustrate the expression profiles of these three variables A, B and C with respect to a 2-phenotype problem (*i.e.*, where there are two different phenotypes) as shown in Fig. 1b. The variable B is having high positive dependence with both A and C in phenotype 1, but negative dependence with these in phenotype 2. It shows that B is differentially co-expressed with A and C, but differential co-expression cannot exist between A and C. In other words, transitive dependence does not hold for differential co-expression. Let us formalize this as the following lemma.

Lemma 3.1 *Let $\mathcal{DC}(x, y)$ denote that two biomolecules x and y are differentially co-expressed in a 2-phenotype problem. Then $\mathcal{DC}(m_2, m_3)$ cannot be true if both the cases $\mathcal{DC}(m_1, m_2)$ and $\mathcal{DC}(m_1, m_3)$ hold.*

In network-based analyses, it is instrumental to explore sets (of cardinality more than two) of strongly connected biomolecules. But, for a network of differentially co-expressed biomolecules, it is hard to characterize a dense module due to the nonexistence of transitivity (as observed in Lemma 3.1). So, clustering based approaches become less efficient while finding a set of differentially co-expressed miRNAs from a network. Based on this observation, we restrict our approach to pairwise dependence analysis of miRNAs rather than finding modules (or groups). We introduce the following terminology to discuss in detail about some novel properties of differential co-expression networks.

Definition 3.1 (Switching pattern) *A switching pattern of a pair of differentially co-expressed biomolecules, denoted as $(p_1\langle d\rangle, p_2\langle d'\rangle)$, defines the differential pattern of dependence within a pair of phenotypes p_1 and p_2, where $\langle d\rangle$ represents either positive or negative dependence and $\langle d'\rangle$ is its complement.*

Note that the *switching pattern* can be of two types – (p_1+, p_2-) and (p_1-, p_2+), both of which symbolize differential patterns for a pair of phenotypes p_1 and p_2. In light of the current analysis, the types of dependence denote co-expression and contra-expression. Using the notion of *switching pattern*, we formally define a differentially co-expressed network as follows.

Definition 3.2 (Differentially Co-expressed Network) *A differentially co-expressed network, for a 2-phenotype problem with phenotypes p_1 and p_2, is defined as a quadruplet $G = (V, E, W, S)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is a set of nodes denoting biomolecules, $E \subseteq V \times V$ is the set of edges between pairs of biomolecules that are differentially co-expressed, $W : E \rightarrow R^+$ is a weight function representing the strength of differential co-expression, and $S : E \rightarrow \{(p_1+, p_2-), (p_1-, p_2+)\}$ denotes the switching pattern between the biomolecules.*

The characteristics of a differentially co-expressed network G are given below:

- G is not necessarily connected (G might contain more than one component)
- There is no isolated node in G (differential co-expression is always defined on an edge)
- G is undirected (the weights and *switching patterns* associated with the edges (v_i, v_j) and (v_j, v_i) are indistinguishable).
- G is acyclic (there is no cycle due to nonexistence of transitivity)

We are interested to explore the substructures, which have same differential patterns, in the acyclic differentially co-expressed networks. As transitive dependency does not hold also for such substructures they are essentially trees. We formalize such a structure as a differentially co-expressed switching tree, hereafter abbreviated as DCST. In general, it is a weighted and undirected network without cycles having edges between node pairs denoting

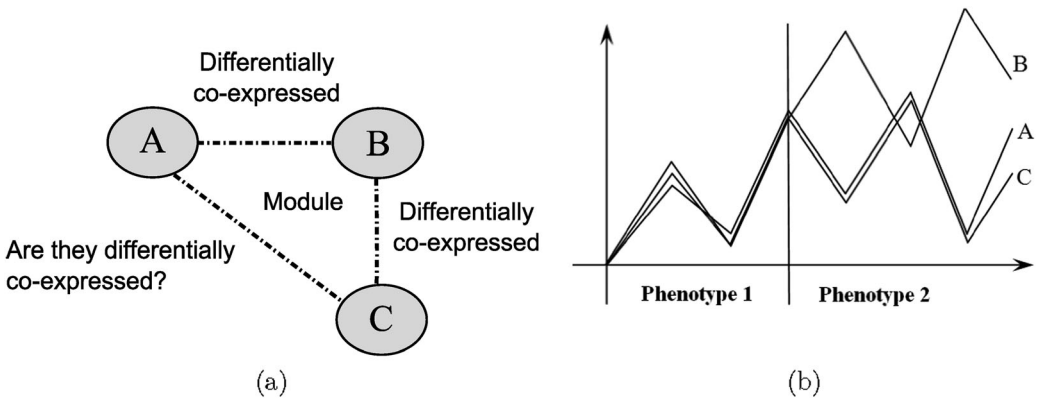

Fig. 1 (a) Possibility of the formation of differentially co-expressed module between the biomolecules A, B and C. (b) Differential co-expression patterns between A, B and C. Time points are taken along the *X*-axis and *Y*-axis denotes the expression profiles.

differentially co-expressed biomolecules. The only constraint is that the same *switching pattern* appears throughout the DCST. The formal definition of a DCST is as follows.

Definition 3.3 (DCST) *Given a differentially co-expressed network $G = (V, E, W, S)$, a DCST is defined as a subnetwork $G' \subseteq G$ having the same switching pattern over all the edges.*

above algorithm. Note that there exists only two types of *switching patterns* for a 2-phenotype problem, namely (p_1+, p_2-) and (p_1-, p_2+). T is selected in such a way that ensures a valid *switching pattern* in the DCST. Therefore, the choice of the parameter T is an important performance factor of the algorithm. The time requirements of Algorithm 1 is derived below.

Algorithm 1 Finding the large DCSTs in a differentially co-expressed network.

Input: A differentially co-expressed network $G = (V, E, W, S)$, a threshold of the differential co-expression weight T, and a degree threshold T_D.
Data Structure: An edge listing of G along with the edge weights and *switching patterns*.
Output: The set of large DCSTs $\{G_1 = (V_1, E_1, W_1, S_1), G_2 = (V_2, E_2, W_2, S_2), \ldots, G_n = (V_n, E_n, W_n, S_n)\}$.
Initialization: $G_1 = G_2 = \ldots = G_n = \phi$ and $k = 0$.
Algorithmic Steps:

1: $E \leftarrow E - (v_i, v_j), \forall v_i, v_j \in V$ and $W_{v_i v_j} \leq T$.
2: **for** each $v_i \in V$ **do**
3: $d(v_i) = \#(v_i, v_j), \forall v_j \in V - \{v_i\}$ and $(v_i, v_j) \in E$.
4: **end for**
5: **repeat**
6: $k \leftarrow k + 1$.
7: Find the seed edge $(v_i^s, v_j^s) = \arg_{(v_i, v_j)} \max_{\forall v_i, v_j \in V} W_{v_i v_j}$.
8: Include (v_i^s, v_j^s) and the corresponding vertices, edge weight and the pattern of switching in G_k.
9: **repeat**
10: Select the node $v^* \in V - V_k$, by resolving ties arbitrarily, to expand the current DCST G_k further by the inclusion of the edge (v_i, v^*) such that:
a. $v_i \in V_k$.
b. $W_{v_i v^*} = \max_{\forall v_j \in V - V_k} W_{v_j v^*}$.
c. $d(v^*) > T_D$.
d. $S_{v_i v^*}$ equals $S_{v_i^s v_j^s}$.
11: **until** The current DCST G_k is no more expandable by the inclusion of edges
12: **if** G_k is empty **then**
13: Exit.
14: **else**
15: Return G_k and remove the nodes V_k and associated edges from G.
16: **end if**
17: **until** The reduced network is empty

To find out the set of large DCSTs from a given differentially co-expressed network, constructed upon pairwise dependence between biomolecules, we propose a deterministic algorithm. The approach is formally presented as Algorithm 1. For brevity, let us assume that $d(v_i)$ denotes the degree (*i.e.*, the number of other nodes to which it is connected) of a node v_i, and the weight and *switching pattern* of an edge (v_i, v_j) is denoted as $W_{v_i v_j}$ and $S_{v_i v_j}$, respectively.

In the beginning (step 1) of the algorithm, the given network is filtered such that only the edges with high differential co-expression values are retained (*i.e.*, those that are above a threshold T). In this way, we can reduce the weighted network by removing poorly differentially co-expressed node pairs. Then (steps 2–4) the degree values of the nodes are computed. Following this, an iterative process is pursued (steps 5–17) to prepare a complete set of DCSTs. For this, initially a seed edge, the one with maximum differential co-expression weight, is selected. Then, in the inner part of the algorithm (steps 9–11), the edges are expanded starting from the seed edge by including edges of the remaining network that have maximum differential co-expression with respect to the nodes in the current DCST and high degree (steps 10(a–d)). Pursuing the same process, we can finally generate a set of DCSTs using the

Lemma 3.2 *The worst case time complexity of Algorithm 1 is $O(n^2)$ and space complexity is $O(n^2)$, where n is the number of nodes in the network.*

Proof The initial filtering step of Algorithm 1 takes at most $O(n^2)$ time to verify the weights of all edges. This step ensures the non-existence of cycles in G. Therefore, the number of edges modifies into the order of the nodes. Following this, the degree computation can be done in $O(n^2)$ time for all the n nodes. Then in each iteration, the selection of the edge with maximum weight (seed edge) requires $O(n)$ time. Once the seed edge is selected, the process of expanding the seed edge might take – $O(n)$ time to find out the node having maximum differential co-expression with the remaining network (this computation for all the nodes needs not to be repeated in each iteration), $O(n)$ time for identifying the maximum degree, and $O(1)$ time for verifying the *switching pattern* match. In the worst case, extracted DCST will contain only the initialized seed edge. So, the entire process might iterate $O(n)$ times. Thus, the total worst case time complexity becomes:

$$O(n^2) + O(n^2) + O(n) \times (O(n) + O(n) + O(1)) = O(n^2).$$

The space complexity is obviously to keep the edge listing with weights and *switching patterns* in $O(n^2)$ space.

Therefore, the sub-optimal approach provided in Algorithm 1 requires time and space in quadratic order. It is thus effective for larger networks and scalable for the problems like studying the differential co-expression networks of genes or miRNAs.

4 Experimental analysis

Results are provided on an expression dataset of 170 miRNAs profiled separately in GM and WM for several early AD subjects using Exiqon Locked Nucleic Acid (LNA) microarrays.[15] The details about the data, along with the details about expression profiling and study groups, can be found in section 5 of ESI.† From this dataset, first the differential co-expression network is built and the DCSTs are identified. Then we describe how these DCSTs can be utilized to gain novel and significant biological insights. We observe that the miRNA miR-423-5p acts as a hub in the network. In course of the current study, we show that we can identify the miRNAs associated with AD using our approach. To evaluate the effectiveness of this, as well as some of the existing similar ones, we prepare a list of miRNAs that have already been validated in experimental studies to have a role in AD. These are reported together in the ESI† (see Table S1, ESI†). We refer to this set as the Gold standard information. It contains 74 distinct miRNAs in total. Based on this, we evaluate how better we can identify the AD involvement of miRNAs. We also point out the significant involvement of WM in early AD progression, a fact less acknowledged earlier.

4.1 Analysis of the DCSTs

In this part, we describe how to build the differential co-expression network, find DCSTs from it, and further analyze their significance, statistically and from the perspective of Alzheimer's progression.

4.1.1 Building up the differential co-expression network. Correlation is possibly the best measure for co-expression analysis because it can simultaneously model positive and negative dependence (albeit linear) between a pair of variables.[25] We compute the Pearson correlation coefficient (ρ) of expression values between the miRNA pairs (see section 6 in ESI† for more details) separately in GM ($\rho_{GM} \in [-1, 1]$) and in WM ($\rho_{WM} \in [-1, 1]$). Then we calculate the absolute difference of correlation (ADC) between these two. This ADC value is used to designate differentially co-expressed miRNA pairs. Those miRNA pairs having ADC value greater than 1 are considered as the differentially co-expressed miRNAs. This is because the inequality $|\rho_{GM} - \rho_{WM}| > 1$ implies either $\rho_{GM} \in (0, 1]$ and $\rho_{WM} \in [-1, 0)$ or $\rho_{GM} \in [-1, 0)$ and $\rho_{WM} \in (0, 1]$. Therefore, by setting a threshold value of ADC > 1, we are sure to obtain differentially co-expressed miRNAs such that if a pair of miRNAs is co-expressed (or contra-expressed) in GM, then respectively they must be contra-expressed (co-expressed) in WM. The ADC value is considered here as the degree of differential co-expression between an miRNA pair. We use this ADC value (>1) as the weight function (W) to construct differentially co-expressed networks for further analysis.

To verify how relevant the differentially co-expressed miRNAs identified by the ADC values are, the top 15 miRNA pairs having the highest ADC values are extracted from the dataset (listed in Table S2(a), ESI†). The top 15 pairs resulted in a set of 16 miRNAs. We found that 10 out of the 16 miRNAs (p-value = 4.1×10^{-3}) thus obtained match with the Gold standard list of AD related miRNAs. We further observe that the miR-423-5p appears in all the 15 miRNA pairs selected based on high ADC values (the degrees of connectivity between these miRNA pairs are shown in Table S2(b), ESI†). It is already known that this miR-423-5p is associated with AD progression,[14] and its significant differential co-expressibility (as we observed) with the others inevitably highlight it as a hub miRNA. It is also evident from Table S2 (ESI†) that the top 15 miRNA pairs show highly negative correlation in GM and positive correlation in WM (see section 6 of ESI†). So, we do not obtain separate sets of top 15 miRNAs differentially co-expressed in GM and WM.

4.1.2 Emergence of miR-423-5p as a hub. The presence of miR-423-5p in all the miRNA pairs found in the correlation-based analysis inevitably raises the question whether it is an outlier. The outlier is an observation that numerically deviates markedly from the other observations of the sample from which it is taken. It is possible that the expression values profiled for miR-423-5p may have experimental errors. To verify this, the correlation-based analysis is repeated (*i.e.*, the ADC values between miRNA pairs in GM and WM is computed) by excluding miR-423-5p when 17 miRNAs are obtained in the top 15 miRNA pairs. The results are provided in Table S3 (ESI†). As can be seen from the table, with this approach a smaller number of validated AD-associated miRNAs are uncovered, in comparison with the earlier results shown in Table S2 (ESI†). Only 4 miRNAs out of 17 (p-value = 7.97×10^{-1}) are found to be AD-associated as evaluated from the Gold standard (results reported in Table S1, ESI†). It confirms that instead of being an outlier, miR-423-5p is possibly working as a hub in the interacting network between the differentially expressed miRNAs related to AD. This miRNA is thus important for further study.

We also analyze the significance of the miRNAs not selected by the correlation-based analysis, but present in the Gold standard list. By pursuing degree-based analysis on this, we highlight that there is equal importance of the degree of differential co-expression and also the degree of connectivity (further details in ESI† section 6). So, we keep a degree threshold in our proposed approach of exploring the DCSTs.

4.1.3 Applying the proposed algorithm. We apply the proposed methodology given in Algorithm 1 to the differential co-expression network of miRNAs obtained by assuming the ADC values as weights. The weight threshold is taken as $T = 1$ (as discussed in section 4.1.1). To select the degree threshold of T_D, we have analyzed the degree distribution of the miRNAs obtained from the constructed network. The frequency distribution of the degree values is shown in Fig. 2. A peak can be observed in this plot around a degree value of 30 which leads us to the consideration of this value as the degree threshold.

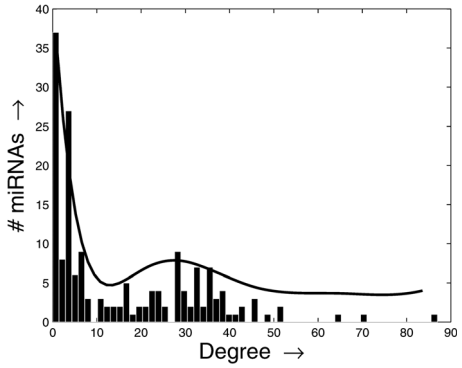

Fig. 2 Frequency distribution of degrees of miRNAs.

The modules obtained as DCSTs are shown in Fig. 3. Interestingly, we obtain two separate DCSTs with two different *switching patterns*. It can be observed from the DCST (say DCST 1) shown in Fig. 3a that all the miRNAs have negative co-expression in GM and positive co-expression in WM, whereas the reverse pattern is observed in the other DCST, say DCST 2, shown in Fig. 3b. DCTS 1 has 10 miRNAs and DCST 2 has 21 miRNAs. These two tree substructures are further evaluated to verify how much AD-associated they are.

We report about the AD-association of all the miRNAs, obtained in the DCSTs, in Table 1. Table 1a and b show the type of enrichment and validated association of miRNAs with AD. We note that 6 out of 10 miRNAs in DCST 1, and 8 out of 21 miRNAs in DCST 2, are present in the Gold standard list of AD-association, even though their *switching patterns* are completely opposite. So, both of these patterns can model dysregulation due to disease participation of miRNAs. Interestingly, we observe that almost all the miRNAs (11 out of 14, p-value < 0.01) found to be related with AD are enriched in WM. It is also important that the miRNAs having positive co-expression in WM have higher co-expression values than the miRNAs having positive co-expression in GM. From these observations, we are able to infer that high differential co-expression of miRNAs in WM may be highly related to AD pathology.

4.1.4 Statistical significance analysis of the DCSTs. To further verify whether the DCSTs we obtain with the proposed approach are significant substructures, we pursue a randomization test on the differential co-expression network. The purpose is to see the effects of perturbation on the network by randomization. We perturb the network and find the occurrence of DCSTs in that (more details can be found in ESI† section 7). It turns out that the sizes of the DCSTs we obtain for two different *switching patterns* are statistically significant (p-value < 0.001).

4.1.5 Biological significance of DCSTs. On further examining the regulatory information about the miRNAs found in the DCSTs (total 31 in two sets) from recent resources,[26,27] we observe that many of the regulatory TFs of these miRNAs are common. For example, TFs like EGR1 and E2F1 are common regulators of many miRNAs as per TransmiR.[27] Very recently, it has been identified that EGR1 is up-regulated by γ-secretase inhibitor, an important molecular drug target for AD, in human osteosarcoma cell lines.[28] There are also early evidences of the regulatory activities of EGR1 on the AD-associated gene presenilin-2 (PSEN2) and its implications in neuronal plasticity.[29] The other transcriptional regulator E2F1 is a brain-specific protein in humans having reported association with the AD.[30,31] There is another recently published web server mirGen 2.0[26] that reports the TFs binding to the promoter region of many miRNAs. While analyzing this data, we found several common TFs such as Pax-4, Oct-1, Evi-1, HNF-4, *etc.* regulating the miRNAs appearing in the DCSTs. Thus, these TFs may be significant in the Alzheimer's pathway in humans and should be looked into further.

4.1.6 Involvement of WM in the progression of AD. After careful literature survey, we find that WM has been rarely hypothesized or shown to have a significant role in AD. The WM is already known to have significant role in managing the communication and connectivity in the brain. Therefore, the loss of WM might retard the speed of information processing and develop the chance of forgetfulness.[32] Roher *et al.* are the first to claim that AD may also originate in the WM of brain, showing that alterations in the WM may cause changes in GM.[33] This is in contrast to the previous belief that the GM is first affected in AD, which then propagates to the WM. Additionally, they have highlighted that WM alterations in AD cases are predominantly found in women than in men. This has also been verified in a recent study on high-AD-risk groups of women using multimodal imaging.[34] Remarkably, the expression profiles that we consider in this paper are also taken from a study on elderly women.[15] Therefore, early Alzheimer's progression in elderly women seems to be associated with some activities of miRNAs (and associated TFs and genes) in the WM. However, it will be hard to make concrete decisions based on this. It might require sex-specific and independent further studies to get confirmation on this. Either we require more specific experimental studies to nullify this initial belief or otherwise accept this and look more into the reason. Nevertheless, *in vitro* miRNA activity in WM appears to have some important relation with AD progression and WM lesions are already known to be risk factors for AD.[35]

As a whole, we become more sure that the observation of WM-enrichment of the miRNAs in the DCSTs is biologically significant. Interestingly, all these miRNAs are obtained from two separate DCSTs that have opposite *switching patterns*, even though their enrichment types are same. So, this observation points to the fact that the role of WM in AD progression is indeed crucial, and might have been overlooked in the earlier studies.

4.2 Analysis with existing methods

In course of our current analysis, we are able to identify AD-associated miRNAs from differential co-expression networks. This is accomplished by mining special substructures (DCSTs) in these networks. Conventional graph clustering methods can also find substructures from a network. Therefore, we evaluated some graph clustering algorithms to verify how well they can find AD-associated miRNAs. Methods based

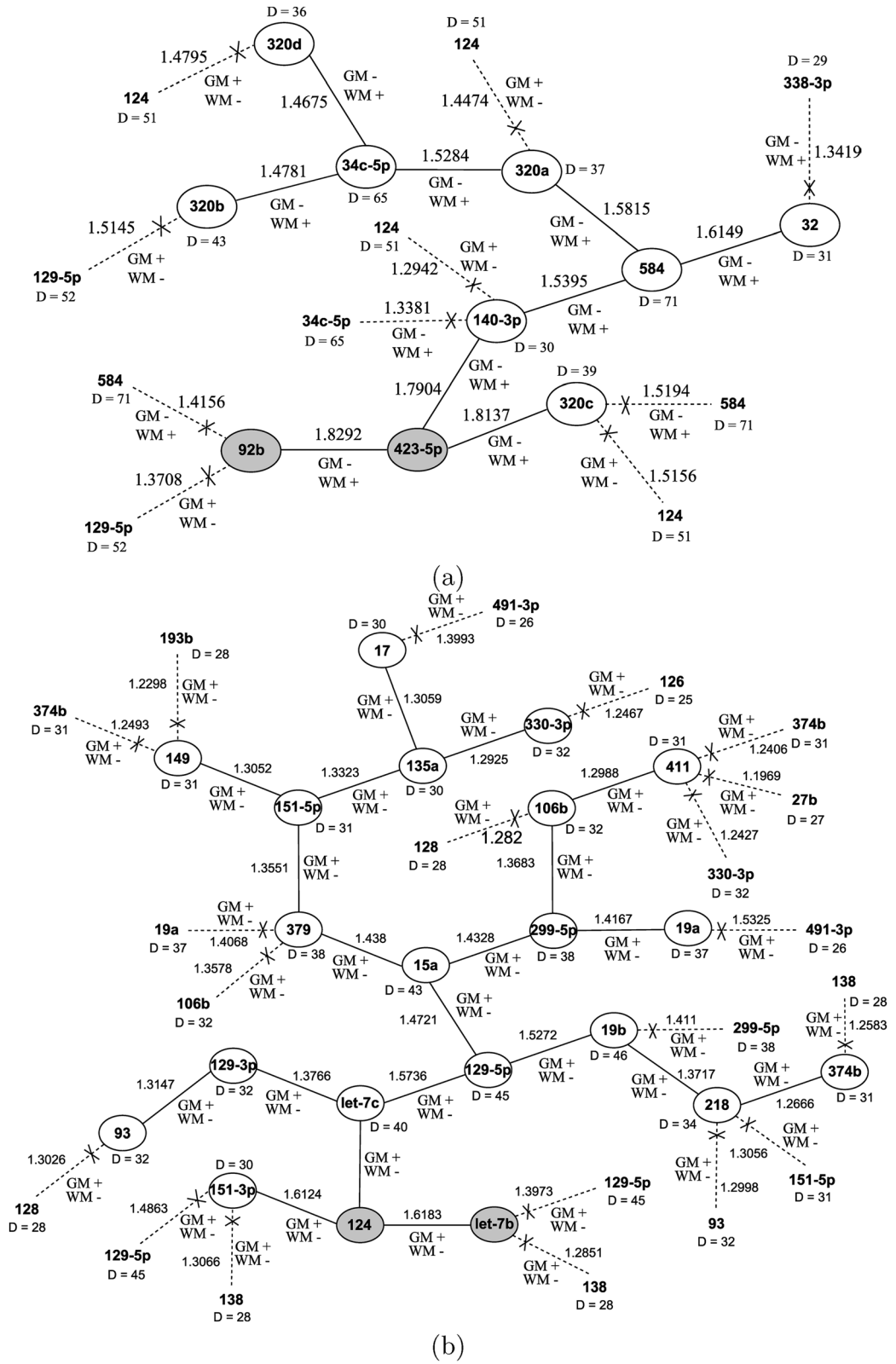

Fig. 3 The DCSTs with nodes labeled with the identifiers of human miRNAs. The gray ellipses denote the nodes forming the seed edge and D stands for the degree value. (a) DCST 1 with 10 miRNAs which are negatively co-expressed in GM and positively co-expressed in WM. (b) DCST 2 with 21 miRNAs which are positively co-expressed in GM and negatively co-expressed in WM.

on differential expression are also useful for this purpose. So, we adopted some analyses in light of such existing approaches. In this section, we apply different graph clustering and differential expression based tools to check which can uncover the

Table 1 The enrichment of the miRNAs obtained in (a) DCST 1 with the (GM+, WM−) pattern and (b) DCST 2 with the (GM−, WM+) pattern

	Associated with AD		Not associated with AD	
miRNA name	Enriched in WM	Enriched in GM	Enriched in WM	Enriched in GM
(a)				
miR-92b	√			
miR-423-5p	√			
miR-320c	√			
miR-140-3p			√	
miR-584			√	
miR-32			√	
miR-320a	√			
miR-34c-5p			√	
miR-320b	√			
miR-320d	√			
Total	6	0	4	0
(b)				
miR-124		√		
let-7b				√
miR-151-3p				√
let-7c				√
miR-93	√			
miR-129-3p		√		
miR-129-5p		√		
miR-19b	√			
miR-218				√
miR-374b			√	
miR-15a	√			
miR-379				√
miR-299-5p				√
miR-19a			√	
miR-151-5p			√	
miR-106b	√			
miR-149				√
miR-135a				√
miR-330-3p			√	
miR-411				√
miR-17	√			
Total	5	3	4	9

maximum number of miRNAs validated to have association with AD. The following subsections provide the details of the studies carried out.

4.2.1 Graph clustering. Clustering of the differentially co-expressed network could be an approach for identifying AD related miRNAs. We apply some graph-theoretic approaches in this subsection to find out significant miRNAs associated with AD. Initially, we construct an unweighted network with miRNAs taken as nodes and having ADC values higher than 1 as edges. After this, we prepare the adjacency matrix corresponding to this unweighted network. Finally, several graph clustering algorithms are applied (detailed in ESI† section 10) on this unweighted network to find out highly interacting node sets (miRNAs). We use two separate measures to find out the density (how strongly interconnected) and biological importance (how many are associated with AD) of the graph clusters. But no dependence is observed between these two measures. As we pointed out earlier (section 3), extracting modules by grouping miRNAs of similar differential co-expression patterns is not a valid approach, which is the basic motivation of graph clustering algorithms. Therefore, tree substructures (without cycles), in the form of DCSTs are possibly the best representative set of disease-miRNAs instead of the modules.

4.2.2 Paired student's *t*-test. Wang *et al.* have used the *t*-test on their own experimental data to find out the miRNAs enriched in GM and WM.[15] We replicate the same analysis using paired student's *t*-test (details can be found in ESI† section 8). The top 15 miRNAs with lowest *p*-values, differentially expressed either in GM or in WM, are selected. As we found, 8 miRNAs in GM (p-value = 3.74×10^{-2}) and 4 miRNAs in WM (p-value = 6.96×10^{-1}) out of the top 15 match with the Gold standard list of AD related miRNAs (vide Table S7, ESI†). The *p*-values are quite high indicating that the result is not significant. So, this type of approach with differential expression may not be suitable for the analyzed data. The inefficacy of the *t*-test might be justified by the small sample size used in the current analysis (see section 8 in ESI†).

4.2.3 Significance analysis of microarrays. To further verify the strength of differential expression analysis on the data,[15] we apply the significance analysis of microarrays (SAM) method[5] on the expression profiles of all the 170 miRNAs. The output of SAM provides a list of statistically significant set of miRNAs (see section 9 of ESI†). We obtain two sets of miRNAs – one having higher expression in WM and the other in GM. We extract the top 15 miRNAs from both these sets. We observe a match of 5 miRNAs in GM (p-value = 4.66×10^{-1}) and 4 miRNAs in WM (p-value = 6.96×10^{-1}) with the Gold standard (provided in Table S8, ESI†) out of the top 15 selected in both cases. As in the case of paired student's *t*-test, here also the *p*-values are very high, further strengthening that differential expression analysis may not be a good tool for identifying AD related miRNAs from this data.

To further verify how SAM ranks the miRNAs that are listed in the Gold standard but are not selected in the top 15 lists obtained for GM and WM, we study their *p*-values (Table S9 in ESI†). Evidently, the miRNAs belonging to the Gold standard list are found to have good ranks by the analysis with SAM (more details in the ESI† section 9), though they are filtered out with the stringent criterion of selecting only the top 15 miRNAs. However, several AD associated miRNAs could be identified using the differential co-expression analysis. This underlines the importance of differential co-expression analysis over differential expression analysis for studies related to AD. The reason may be the complex functional behavior of miRNAs present in the human brain. That is why, we chose the approach of differential co-expression analysis for the current analysis, based on which we construct the differentially co-expressed network and explore the DCSTs.

5 Concluding remarks

In this paper, we have first surveyed, in detail, the miRNAs, their expression, and their association with AD pathology. Thereafter, we prepare a list of AD related miRNAs through extensive literature survey. The main focus of this study is to find out the role of miRNAs in AD progression based on the switching of their differential co-expression pattern.

The preliminary motivation is to find out AD-associated miRNAs from expression profiles. We show through several analyses that the existing measures and methodologies are incapable of fulfilling the motivation. Through these analyses, we learn some important ideas which are used further to build up a specialized algorithm for finding out significant substructures from the network of interacting miRNAs. We focus on mining a kind of substructures, called DCSTs, which contain a set of miRNAs forming a tree having similar type of *switching patterns* of differential co-expression. The set of miRNAs emerging through this substructure mining contains a significant number of WM-enriched miRNAs. By further study, we confirm a significant role of WM in the progression of AD, an observation that has not been investigated in-depth till date. Thus, the region-specific miRNA expression profiling appears to be an important direction of study in Alzheimer's research.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments that greatly helped to improve the quality of the paper.

References

1 D. P. Bartel, MicroRNAs: target recognition and regulatory functions, *Cell*, 2009, **136**(2), 215–233.
2 N. N. Hooten, K. Abdelmohsen, M. Gorospe, N. Ejiogu, A. B. Zonderman and M. K. Evans, MicroRNA expression patterns reveal differential expression of target genes with age, *PLoS One*, 2010, **5**(5), e10724.
3 W. X. Wang, B. W. Rajeev, A. J. Stromberg, N. Ren, G. Tang, Q. Huang, I. Rigoutsos and P. T. Nelson, The expression of microRNA miR-107 decreases early in alzheimer's disease and may accelerate disease progression through regulation of beta-site amyloid precursor protein-cleaving enzyme 1, *J. Neurosci.*, 2008, **28**(5), 1213–1223.
4 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*, 2010, **5**(6), e11070.
5 V. G. Tusher, R. Tibshirani and G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**(9), 5116–5121.
6 N. Filigheddu, I. Gregnanin, P. E. Porporato, D. Surico, B. Perego, L. Galli, C. Patrignani, A. Graziani and N. Surico, Differential expression of microRNAs between eutopic and ectopic endometrium in ovarian endometriosis, *J. Biomed. Biotechnol.*, 2010, 369–549.
7 Z. G. Mao, D. S. He, J. Zhou, B. Yao, W. W. Xiao, C. H. Chen, Y. H. Zhu and H. J. Wang, Differential expression of microRNAs in GH-secreting pituitary adenomas, *Diagn. Pathol.*, 2010, **5**(1), 79.
8 S. Ura, M. Honda, T. Yamashita, T. Ueda, H. Takatori, R. Nishino, H. Sunakozaka, Y. Sakai, K. Horimoto and S. Kaneko, Differential microRNA expression between hepatitis B and hepatitis C leading disease progression to hepatocellular carcinoma, *Hepatology*, 2009, **49**(4), 1098–1112.
9 C. F. Stählera, A. Kellera, P. Leidingerb, C. Backesb, A. Chandranc, J. Wischhusenc, B. Meder and E. Meese, Whole mirnome-wide differential co-expression of microRNAs, *Genomics, Proteomics Bioinf.*, 2012, **10**(5), 285–294.
10 M. C. Oldham, S. Horvath and D. H. Geschwind, Conservation and evolution of gene coexpression networks in human and chimpanzee brains, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**(47), 17973–17978.
11 K. Duff and F. Suleman, Transgenic mouse models of alzheimer's disease: How useful have they been for therapeutic development?, *Briefings Funct. Genomics Proteomics*, 2004, **3**(1), 47–59.
12 J. A. Saugstad, MicroRNAs as effectors of brain function with roles in ischemia and injury, neuroprotection, and neurodegeneration, *J. Cereb. Blood Flow Metab.*, 2010, **30**(9), 1564–1576.
13 M. A. Christensen, W. Zhou, H. Qing, A. Lehman, S. Philipsen and W. Song, Transcriptional regulation of BACE1, the beta-amyloid precursor protein beta-secretase, *Mol. Cell. Biol.*, 2003, **24**(2), 865–874, by Sp1.
14 J. Satoh, MicroRNAs and their therapeutic potential for human diseases: Aberrant microRNA expression in Alzheimer's disease brains, *J. Pharmacol. Sci.*, 2010, **114**(3), 269–275.
15 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 Neuropathol.*, 2010, **121**(2), 193–205.
16 L. M. Ittner, N. Schonrock, M. Matamales and J. Götz, MicroRNA networks surrounding APP and amyloid-β metabolism – implications for alzheimer's disease, *Exp. Neurol.*, 2012, **235**(2), 447–454.
17 J. Y. Yu, K. H. Chung, M. Deo, R. C. Thompson and D. L. Turner, MicroRNA miR-124 regulates neurite outgrowth during neuronal differentiation, *Exp. Cell Res.*, 2008, **314**(14), 2618–2633.
18 R. B. Campenot, Local control of neurite development by nerve growth factor, *Proc. Natl. Acad. Sci. U. S. A.*, 1977, **74**(10), 4516–4519.
19 J. K. Choi, U. Yu, O. J. Yoo and S. Kim, Differential coexpression analysis using microarray data and its application to human cancer, *Bioinformatics*, 2005, **21**(24), 4348–4355.
20 M. Watson, Coxpress: differential co-expression in gene expression data, *BMC Bioinformatics*, 2006, **7**, 509.
21 J. Gillis and P. Pavlidis, A methodology for the analysis of differential coexpression across the human lifespan, *BMC Bioinformatics*, 2009, **10**, 306.
22 Y. Choi and C. Kendziorski, Statistical methods for gene set co-expression analysis, *Bioinformatics*, 2009, **25**(21), 2780–2786.

23 G. Fang, R. Kuang, G. Pandey, M. Steinbach, C. L. Myers and V. Kumar, Subspace differential coexpression analysis: Problem definition and a general approach, in *Proceedings of the Pacific Symposium on Biocomputing*, 2010, vol. 15, pp. 145–156.
24 T. Idekera and N. J. Krogan, Differential network biology, *Mol. Syst. Biol.*, 2012, **8**, 565.
25 S. Bandyopadhyay and M. Bhattacharyya, A biologically inspired measure for co-expression analysis, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2011, **8**(4), 929–942.
26 P. Alexiou, T. Vergoulis, M. Gleditzsch, G. Prekas, T. Dalamagas, M. Megraw, I. Grosse, T. Sellis and A. G. Hatzigeorgiou, miRGen 2.0: a database of microRNA genomic information and regulation, *Nucleic Acids Res.*, 2009, **38**, D137–D141.
27 J. Wang, M. Lu, C. Qiu and Q. Cui, TransmiR: a transcription factor-microRNA regulation database, *Nucleic Acids Res.*, 2009, **38**, D119–D122.
28 Y. Matsunoshita, K. Ijiri, Y. Ishidou, S. Nagano, T. Yamamoto, H. Nagao, S. Komiya and T. Setoguchi, Suppression of osteosarcoma cell invasion by chemotherapy is mediated by urokinase plasminogen activator activity *via* up-regulation of EGR1, *PLoS One*, 2011, **6**(1), e16234.
29 P. Renbaum, R. Beeri, E. Gabai, M. Amiel, M. Gal, M. U. Ehrengruber and E. Levy-Lahad, Egr-1 upregulates the Alzheimer's disease presenilin-2 gene in neuronal cells, *Gene*, 2003, **318**, 113–124.
30 A. Giovanni, E. Keramaris, E. J. Morris, S. T. Hou, M. O'Hare, N. Dyson, G. S. Robertson, R. S. Slack and D. S. Park, E2F1 mediates death of B-amyloid-treated cortical neurons in a manner independent of p53 and dependent on Bax and Caspase 3, *J. Biol. Chem.*, 2000, **275**, 11553–11560.
31 J. Rhodes, K. Jordan-Sciutto and R. Bowser, Altered subcellular distribution of transcriptional regulators in response to Aβ peptide and during Alzheimer's disease, *Mech. Ageing Dev.*, 2001, **123**(1), 11–20.
32 B. Pakkenberg and H. J. G. Gundersen, Neocortical neuron number in humans: effect of sex and age, *J. Comp. Neurol.*, 1997, **384**, 312–320.
33 A. E. Roher, N. Weiss, T. A. Kokjohn, Y. Kuo, W. Kalback, J. Anthony, D. Watson, D. C. Luehrs, L. Sue, D. Walker, M. Emmerling, W. Goux and T. Beach, Increased Aβ peptides and reduced cholesterol and myelin proteins characterize white matter degeneration in alzheimer's disease, *Biochemistry*, 2002, **41**(37), 11080–11090.
34 B. T. Gold, D. K. Powell, A. H. Andersen and C. D. Smith, Alterations in multiple measures of white matter integrity in normal women at high risk for alzheimer's disease, *NeuroImage*, 2010, **52**(4), 1487–1494.
35 F.-E. de Leeuw, F. Barkhof and P. Scheltens, Progression of cerebral white matter lesions in Alzheimer's disease: a new window for therapy?, *J. Neurol. Neurosurg. Psychiatry*, 2005, **76**(9), 1286–1288.

