RESEARCH

Open Access

Differential co-expression analysis reveals early stage transcriptomic decoupling in alzheimer's disease

Yurika Upadhyaya[1†], Linhui Xie[2†], Paul Salama[2], Sha Cao[3], Kwangsik Nho[4], Andrew J. Saykin[4], Jingwen Yan[1,4*] and for the Alzheimer's Disease Neuroimaging Initiative

Abstract

Background: Alzheimer's disease (AD) is one of the leading causes of death in the US and there is no validated drugs to stop, slow or prevent AD. Despite tremendous effort on biomarker discovery, existing findings are mostly individual biomarkers and provide limited insights into the transcriptomic decoupling underlying AD. We propose to explore the gene co-expression patterns in multiple AD stages, including cognitively normal (CN), early mild cognitive impairment (EMCI), late MCI and AD.

Methods: We modified traiditonal joint graphical lasso to model our asusmption that the co-expression networks in consecutive disease stages are largely similar with critical differences. In addition, we performed subsequent network comparison analysis for identification of stage specific transcriptomic decoupling. We focused our analysis on top AD-enriched pathways.

Results: We observed that 419 edges in CN, 420 edges in EMCI, 381 edges in LMCI and 250 edges in AD were frequently estimated with non zero weights. With modified JGL, the weight of all estimated edges in CN, EMCI and LMCI are zero. In AD group, 299 edges were occasionally estimated to be nonzero and the average correlation between genes was 0.0023. For co-expression change during AD progression, there are 66 pairs of genes that demonstrated a continuously decreasing or increasing co-expression from CN to EMCI, LMCI and AD.The network level clustering coefficient remains stable from CN to LMCI and then decreases significantly when progressing to AD. When evaluating edge level differences, we identified eight gene modules with continuously decreasing or increasing co-expression patterns during AD progression. Five of them shows significant changes from CN to EMCI and thus have the potential to serve system biomarkers for early screening of AD.

Conclusion: We employed a modified joint graphical lasso for estimation of co-expression networks for multiple stages of AD. Comparing with graphical lasso, our modified joint graphical lasso model accounts for the similarity in consecutive disease stages. Our results on real data set revealed five gene clusters with obvious co-expression pattern change from CN to EMCI, which could be used as potential system-level biomarkers for early screening of AD.

Keywords: Gene co-expression network, Stage-specific co-expression changes, Alzheimer's disease

*Correspondence: jingyan@iupui.edu
[†]Yurika Upadhyaya and Linhui Xie contributed equally to this work.
[1]Department of BioHealth Informatics, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA
[4]Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA
Full list of author information is available at the end of the article

Upadhyaya *et al. BMC Medical Genomics* 2020, **13**(Suppl 5):53

Page 2 of 10

Background

Alzheimer's disease (AD) is a major neurodegenerative disorder that has been characterized by gradual memory loss and brain behavior impairment. According to the latest report [1], an estimated number of 5.7 million aging Americans are living with Alzheimer's and this number is expected to escalate in coming years given the rapid increase of aging population. To prevent this public health crisis, tremendous effort has been dedicated to discovery of effective AD biomarkers. In addition to *APOE* e4 alleles known as major genetic determinants [2], large-scale genome-wide association studies (GWAS) have led to identification of many novel genetic risk locus [3]. However, extant work largely investigated genetic variations or individual genes associated with AD. Very few studies paid attention to the interactions and associations among the gene products and how they are gradually disrupted during AD progression [4, 5].

Gene co-expression networks describes the correlation patterns among genes. Differential co-expression analysis examines the altered patterns between co-expression networks of two states, e.g., healthy controls vs. patients. It has great potential to identify the gene clusters affected by stage transition and therefore provide valuable information on how the biological system alters during disease progression. According to the Alzheimer's Disease Neuroimaging Initiative (ADNI) project, subjects usually progress from cognitive normal (CN) status (i.e., those without any signs of depression, dementia or cognitive impairment), to early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI) (i.e., those with deteriorating memory concerns, but not decline in performing daily activities and no signs of dementia), and finally AD. In this project, we propose to perform a differential co-expression analysis across all the stages of AD, including CN, EMCI, LMCI and AD.

One common method to generate co-expression network is through pairwise Pearson's correlation and WGCNA is the widely used package for co-expression analysis [6], which is based on marginal correlation. Although frequently used, these methods do not distinguish direct relationships from indirect relationships, e.g., two genes that show similar co-expression patterns due to a common intermediate gene. Unlike these methods, graphical lasso models the joint distribution of genes and can infer the direct relationships by capturing the conditional independence between genes [7]. A graph generated from graphical lasso, where genes are represented as nodes and their co-expression are represented as edges, provides valuable insights into the transcriptomic coupling under specific condition and therefore can help with generating new biological hypothesis. However, it can only estimate one network at a time. When applied to differential co-expression analysis, they estimate the network for each group separately by treating them as independent. This assumption clearly does not hold in disease research since disease formation is a progressive procedure. Co-expression networks in consecutive disease stages should be largely similar with critical differences. For example, co-expression networks in CN group and EMCI group are expected to be largely similar with critical differences. Toward this, we propose to employ joint graphical Lasso [8] for simultaneous estimation of co-expression networks in multiple disease stages. A joint analysis borrows the strength of the relatedness across disease stages and can potentially reveal the differential co-expression patterns with increased statistical power, which is useful especially when the sample size is very limited.

Given the increasing interest in and evidence of blood-based AD biomarkers [9–12], we focused our analysis on the plasma expression data of genes in top AD-enriched pathways highlighted in [13]. We first estimated the co-expression networks of all diagnosis groups using our modified JGL to better model our assumption of network similarity between consecutive disease stages. Subsequently, we performed a comprehensive network comparison analysis of co-expression networks using edge-level, node-level and network-level metrics. We used global clustering coefficient to identify structural property of each network. Node and edge centrality values were calculated to allow comparison of basic network components present in the network [14] and identification of critical network entities [15]. Finally, we were able to identify eight gene clusters showing gradual changes during the progression of AD. Five of them shows significant changes from CN to EMCI and therefore have the potential to serve as system level biomarkers for early screening of AD.

Methods

Dataset

All the plasma microarray data used in this study were directly obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) project (http://adni.loni.usc.edu). Raw expression values obtained directly from CEL files were pre-processed using the Robust Multi-chip Average (RMA) normalization method. The RMA normalized expression array data further went through several quality control (QC) steps [16]. First, we checked the sex of samples using sex-specific gene expression data, including XIST and USP9Y. Second, sample identity was verified on the basis of expression profiling to Omni2.5M genotype match using a Bayesian method to predict individual SNP genotypes from gene expression data. Detailed data description and the preprocessing steps can be found in ADNI LONI website (http://adni.loni.usc.edu). We focused our analysis on five AD-enriched pathways highlighted in [13]. For each pathway,

Upadhyaya *et al. BMC Medical Genomics* 2020, **13**(Suppl 5):53

Page 3 of 10

we extracted all the involved genes from Metacore. In total, there are 75 genes included in this study (Table 1). In the microarray data, if there are multiple probes corresponding to the same gene, we choose the probe with the maximum mean expression to represent the gene. Gene expression values were adjusted for RNA integrity Number (RIN), baseline age and sex to remove potential bias. Finally, 662 ADNI 1 subjects without missing gene expression values were included. Shown in Table 2 is the detailed demographic information of all subjects. We will jointly learn the gene co-expression networks for four consecutive disease stages.

Joint graphical lasso

We first briefly introduce the graphical lasso method [7]. Suppose we have the expression data of p genes for n subjects in K diagnostic groups. Denote the gene expression profile for k-th group as $\mathbf{X}^{(k)} \subseteq \Re^{n_k \times p}$, where n_k is the number of subjects in k-th group and $k = 1, \ldots, K$. Assuming that the gene expression levels within each group $\mathbf{x}_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)}$ are independent and identically distributed with the positive definite $p \times p$ covariance matrix Σ_k. The values in the covariance matrix reveal how the expression of two genes vary together; the inverse covariance matrix Σ_k^{-1} indicates the conditional independence between pairs of genes. Zero values in the inverse

Table 1 75 genes involved in AD enriched pathways

Pathway	Genes	Total
Immune response-Alternative complement pathway	*C3, CD46, CD55, CD59, CLU, CR1, ITGAM, ITGAX*	8
Development-Neurotrophin family signaling	*MAPK3, BAX, BRAF, GAB1, GRB2, HRAS, JUN, MAGED1, MAP2K1, MAP2K2, MAP3K1, MAPK1,MAPK8, PIK3R1, PIP5K1A, PTPN11, RAC1, RAF1, RAP1A, RAPGEF1, SHC1, SORT1, SOS1, TP53,AK1*	25
Neurophysiological process-NMDA-dependent postsynaptic long-term potentiation in CA1 hippocampal neurons	*AKT1, BRAF, CALM1, CREB1, EIF4E, GRB2, HRAS, MAP2K1, MAP2K2, MAPK1, MAPK3, MKNK1, PIK3R1, PRKACA, RAP1A, RAPGEF3, RPS6KA1, RPS6KB1, SHC1, XYLT2*	20
Cell adhesion-Ephrin signaling	*ADAM10, CDC42, EPHA1, EPHB1, FYN, GRB10, HRAS, KALRN, MAP3K7, MAP4K4, MAPK8, NCK1, NCK2, PAK1, PXN, RAC1, RAF1, RAP1A, RASA1, RHOA, SLA, SLC10A3, SRC, GRB2*	24
Neurophysiological process-nNOS signaling in neuronal synapses	*CALM1, CALM2, CALM3, DLG4, GRIN2C, GRINA, PPP3CA PPP3CB, PPP3CC, PPP3R1, PRKCB, PRKCD, PRKCH, PRKCI, PRKD3*	15

Table 2 Demographic information of participants

Groups	Total(N)	Gender (M/F)	Age (Mean± Std)
CN	225	113/112	76.65±6.16
EMCI	193	105/88	79.26±7.35
LMCI	202	127/75	76.38±7.89
AD	42	27/15	75.69±9.46

covariance matrix Σ_k^{-1} indicates conditionally independence of corresponding gene pairs. That is, two genes are independent of each other after removing the effect of all other genes in the data set.

Traditional graphical lasso estimates individual co-expression networks separately by solving Eq. 1:

$$\min_{\Theta^{(k)}} - \log \det \Theta^{(k)} + trace\left(\mathbf{S}^{(k)}\Theta^{(k)}\right) + \lambda_1 ||\Theta^{(k)}||_1$$

(1)

Where $\mathbf{S}^{(k)} = \left(\mathbf{X}^{(k)}\right)^T \mathbf{X}^{(k)}$ is the empirical covariance matrix. $||\Theta^{(k)}||_1$ is the L_1 norm. $\Theta^{(k)}$ is the inverse covariance matrix (i.e., co-expression network) to be estimated for k-th group. λ_1 is the non-negative tuning parameter to enforce the sparsity of estimated co-expression network.

However, multiple groups may be related and ignoring the common structures shared across groups will inevitably yield sub-optimal results. To address this problem, joint graphical lasso (Eq. 2) was later proposed to enable the estimation of multiple related inverse covariance matrices together through maximizing penalized log likelihood [8]. It borrows the strength of the relatedness across groups and can potentially reveal the differential co-expression patterns with increased statistical power, which is especially useful when the sample size is very limited.

$$\min_{\{\Theta\}} - \sum_{k=1}^{K} n_k \left(\log \det \Theta^{(k)} - trace\left(\mathbf{S}^{(k)}\Theta^{(k)}\right)\right) + P(\{\Theta\})$$
$$s.t. \quad \Theta^{(1)}, \ldots, \Theta^{(K)} > 0$$

(2)

The sparsity within each covariance matrix and the similarity across covariance matrices in K groups are encouraged by penalty $P(\{\Theta\})$. λ_1 and λ_2 are two non-negative parameters to control the enforcement of sparsity and similarity. In [8], it assumed that the co-expression networks across all pair of groups are similar. However, this assumption does not always hold, especially for discovery of disease stage-specific networks. AD is a slowly progressive brain disorder that networks are expected to gradually dissolve or rewire during the progression. Gene co-expression network in the AD patients may have become very different compared to that of cognitive normals after years of progression. Therefore, we modified the penalty term $P(\{\Theta\})$ to Eq. 3 such that the similarity

Upadhyaya *et al. BMC Medical Genomics* 2020, **13**(Suppl 5):53

Page 4 of 10

among networks is only enforced for consecutive stages. For example, networks between CN and EMCI and networks between EMCI and LMCI groups are encouraged to be similar.

$$P\left(\{\mathbf{\Theta}\}\right) = \lambda_2 \sum_{k=1}^{K-1} \sum_{i \neq j} \left|\theta_{ij}^{(k)} - \theta_{ij}^{(k+1)}\right| + \lambda_1 \sum_{k=1}^{K} \sum_{i \neq j} \left|\theta_{ij}^{(k)}\right| \tag{3}$$

To solve the modified JGL model, we followed the steps in [8] using alternating directions method of multipliers (ADMM) algorithm. With the constraints $\mathbf{\Theta}^{(k)} = \mathbf{Z}^{(k)}$, the dual variables $\mathbf{U}^{(k)}$ are introduced to form scaled augmented Lagrangian [17].

$$L\left(\{\mathbf{\Theta}\}, \{\mathbf{Z}\}, \{\mathbf{U}\}\right) = -\sum_{k=1}^{K} n_k \left(\log \det \mathbf{\Theta}^{(k)} - tr\left(\mathbf{S}^{(k)} \mathbf{\Theta}^{(k)}\right)\right)$$
$$+ \frac{\rho}{2} \sum_{k=1}^{K} \left\|\mathbf{\Theta}^{(k)} - \mathbf{Z}^{(k)} + \mathbf{U}^{(k)}\right\|_F^2$$
$$+ P\left(\{\mathbf{Z}\}\right) \tag{4}$$

The exact solution steps can be referred to the JGL paper [8]. We modified the step to update $\{\mathbf{Z}\}$ (Eq. 5).

$$\min_{\{\mathbf{Z}\}} \left[\frac{\rho}{2} \sum_{k=1}^{K} \left\|\mathbf{Z}^{(k)} - \mathbf{A}^{(k)}\right\|_F^2 + P\left(\{\mathbf{Z}\}\right) \right] \tag{5}$$

We found that this minimization problem is completely separable for each element (i,j) in the matrices,

$$\min_{Z_{ij}^{(1)}, \dots, Z_{ij}^{(K)}} \quad \frac{\rho}{2} \sum_{k=1}^{K} \left\|Z_{ij}^{(k)} - A_{ij}^{(k)}\right\|_F^2 + \lambda_1 \sum_{\substack{k=1 \\ i \neq j}}^{K} \left|Z_{ij}^{(k)}\right|$$
$$+ \lambda_2 \sum_{k=1}^{K-1} \left|Z_{ij}^{(k)} - Z_{ij}^{(k+1)}\right| \tag{6}$$

The last penalty term is known as 1-*d* fused lasso and penalizes the absolute differences in adjacent values of β = $[\mathbf{Z}_{ij}^{(1)}, \dots, \mathbf{Z}_{ij}^{(K)}]$. It can be easily reformulated as $\|\mathbf{D}\beta\|_1$, where $\mathbf{D} \subseteq \Re^{(K-1) \times K}$.

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \tag{7}$$

If we first consider λ_1 as zero and set $\mathbf{y} = \left[A_{ij}^{(1)}, \dots, A_{ij}^{(K)}\right]$. The convex optimization problem becomes a simple 1-*d* fused lasso problem (Eq. 8). Note that rank($\mathbf{D}$)= $K - 1$. With a vector $\mathbf{s} \subseteq \Re^K$ that is orthogonal to all the rows in $\mathbf{D}$, we can construct $\hat{\mathbf{D}} = \begin{bmatrix} \mathbf{D} \\ \mathbf{s} \end{bmatrix}$ with rank($\hat{\mathbf{D}}$)=K. Let $\theta = (\theta_1, \theta_2)^T = \hat{\mathbf{D}}\beta$ and $\theta_1 \subseteq \Re^{K-1}$, then Eq. 8 will be

transformed into a regular lasso problem (Eq. 9).

$$\min_{\{\beta\}} \quad \frac{\rho}{2} ||\beta - \mathbf{y}||_F^2 + \lambda_2 \|\mathbf{D}\beta\|_1 \tag{8}$$

$$\min_{\{\theta\}} \quad \frac{\rho}{2} ||\hat{\mathbf{D}}^{-1}\theta - \mathbf{y}||_F^2 + \lambda_2 \|\theta_1\|_1 \tag{9}$$

Note that the L1 penalty is only partially applied to θ_1. In order to solve this, we re-write Eq. 9 to a standard form with transformation $\hat{\mathbf{D}}^{-1}\theta = \left[{}^{K \backslash K-1}X_1 \mid {}^{K \backslash 1}X_2 \right] \left[\theta_1 \mid \theta_2 \right]^T$,

$$\min_{\theta_1, \theta_2} \frac{\rho}{2} \left||(\mathbf{y} - X_1\theta_1) - X_2\theta_2\right||_F^2 + \lambda_2 ||\theta_1||_1 \tag{10}$$

While $\hat{\theta}_2$ can be solved as $\hat{\theta}_2 = \left(X_2^T X_2\right)^{-1} X_2^T \left(\mathbf{y} - X_1\hat{\theta}_1\right)$, the above equation can be rewritten as

$$\min_{\theta_1} \frac{\rho}{2} \left||((\mathbf{I} - \mathbf{P})\mathbf{y} - (\mathbf{I} - \mathbf{P})X_1\theta_1)\right||_F^2 + \lambda_2 ||\theta_1||_1 \tag{11}$$

Where $\mathbf{P} = X_2\left(X_2^T X_2\right)^{-1} X_2^T$. The above equation can be easily solved using the LARS algorithm [18], from which we can back-transform to get the solution for Eq. 8: $\beta = \hat{\mathbf{D}}^{-1}\theta$. By applying soft-thresholding to β, we can get the solution for Eq. 6.

Tuning parameter selection

The tuning parameters λ_1 and λ_2 in the modified JGL model were selected using the Akaike information criterion (AIC) (Eq. 12).

$$AIC(\lambda_1, \lambda_2) = \sum_{k=1}^{K} \left[n_k trace\left(\mathbf{S}^{(k)} \hat{\Theta}_{\lambda_1, \lambda_2}^{(k)}\right) - logdet\hat{\Theta}_{\lambda_1, \lambda_2}^{(k)} + 2E_k \right] \tag{12}$$

Here, $\hat{\Theta}_{\lambda_1, \lambda_2}^{(k)}$ is the estimated inverse covariance matrix for group k under the tuning parameters λ_1 and λ_2. E_k is the number of non-zero elements in $\hat{\Theta}_{\lambda_1, \lambda_2}^{(k)}$. We performed a grid search and λ_1 and λ_2 with minimal AIC score were selected.

Performance evaluation

In the subsequent gene co-expression analysis, we refer our modified JGL model as JGL. To evaluate the performance of JGL, we compared the performance of joint graphical lasso and traditional graphical lasso, where the co-expression network of each group is estimated separately. Using 1000 permuted datasets, we generated 1000 co-expression networks using JGL and derived a frequency network for each group, where each link has a value between 0 and 1000 indicating how many times it is observed to be nonzero. Similarly, we generate another frequency network for each group using the results from graphical lasso. Since all the gene expression has been permuted and they should be independent from each other,

Upadhyaya *et al. BMC Medical Genomics* 2020, **13**(Suppl 5):53

Page 5 of 10

the ground truth value for all edges should be zero. Any nonzero values will be considered as false positives. That is, the value of each edge in the frequency network indicates its chance to be a false positive. We compared the performance of JGL and graphical lasso in terms of how frequently we will observe the false positives using the permuted data set.

Construction of co-expression network
We applied the modified JGL model to estimate the co-expression networks of four diagnostic groups simultaneously. To evaluate the significance of the estimated networks, we generated a random data set by running 1000 times of permutation for each gene and each group. Random data sets were then fed into the modified JGL model to estimate 1000×4 co-expression networks. For each group, we compared our results against these randomly generated networks and estimated an empirical p value for each edge in each group. Edges with empirical P >0.05 were considered insignificant. There were no insignificant edges identified for CN, EMCI and LMCI groups. Five edges in AD were found to be insignificant and were removed from subsequent network comparison analysis.

Centrality of the network nodes
For every node in the network, four weighted centrality values were calculated: weighted degree, weighted betweeness, weighted closeness and weighted clustering coefficient. Weighted degree is the sum of number of connections in terms of the weight of each edge in the network [19]. R package tnet provides the platform to calculate weighted centrality of a network. The weighted degree method proposed method by [20] uses the tuning parameter alpha to tune the connections and connection weights. The alpha value of 1 was used to calculate weighted degree. When an alpha value is 1, it equals to the definition of weighted degree by [19]. That is, degree is based on the connection weights of the network [19, 21]. Similarly, betweenness and closeness centrality are based on the theory of shortest paths in a network. closeness

implies to how close a node is to other nodes in a network. Betweenness is the measure of how often an edge lies in the path of the other edge. We used alpha value as 1 for both calculation. Finally, clustering coefficient explains how connected neighbours of a node are in an overall network. Arithmetic mean measure was used to calculate the global clustering coefficient of the network as well as nodal clustering coefficients [20].

Network comparison
We examined and compared all four co-expression networks using edge level, node level and network level metrics. For each edge, its weight indicates the correlation between two connecting genes. We compared four networks by each edge and looked for edges with continuously increasing or decreasing weights from CN to EMCI, LMCI and AD. Edges with absolute differences >0.01 in consecutive groups (i.e, (CN- EMCI), (EMCI -LMCI) or (LMCI - AD)) were considered as potential biomarkers. We further clustered these edges based on their patterns of change across four disease stages. Similarly, we also compared the centrality values of each node and edge to identify nodes with gradual changes from CN to EMCI, LMCI and AD.

Results
Comparison of jGL and graphical lasso
We examined the difference of our modified JGL and graphical lasso in terms of their performance in estimation of co-expression networks based on permuted data sets. While the permuted data sets are completely random, it is expected that none of the genes are correlated. So all the edges in the estimated networks should have zero weight. However, with networks estimated using graphical lasso, we observed that 419 edges in CN, 420 edges in EMCI, 381 edges in LMCI and 250 edges in AD were frequently (i.e., more than 100 out of 1000 times) estimated with non zero weights (i.e., false positives) (Fig. 1). With modified JGL, the weight of all estimated edges in CN, EMCI and LMCI are zero. In AD group, 299 edges were occasionally (i.e., less than 100 out of
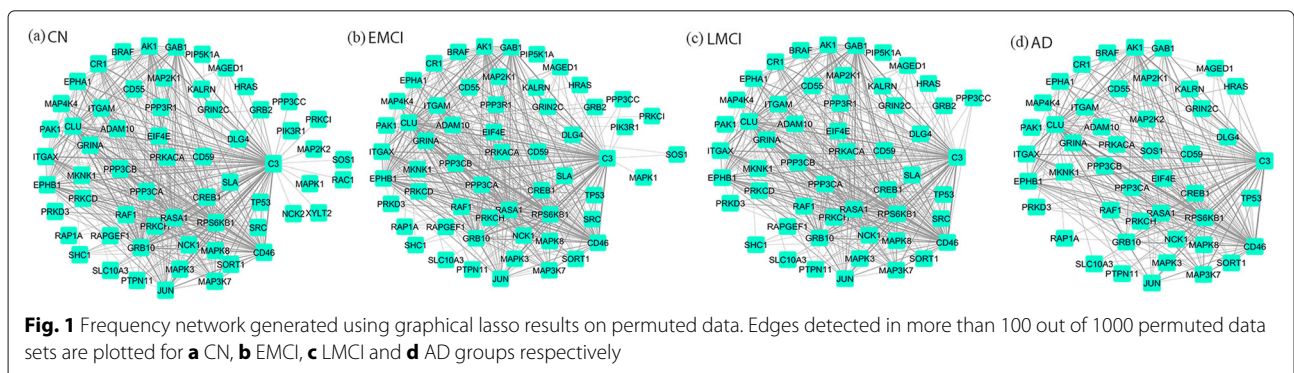
Fig. 1 Frequency network generated using graphical lasso results on permuted data. Edges detected in more than 100 out of 1000 permuted data sets are plotted for **a** CN, **b** EMCI, **c** LMCI and **d** AD groups respectively

Upadhyaya *et al. BMC Medical Genomics* 2020, **13**(Suppl 5):53

Page 6 of 10

Table 3 List of genes with continuous nodal centrality change

Centrality	Type	Nodes
ClusCoef	I	C3, MAPK1, PAK1, PRKCH, SLA
	D	CALM3, MAPK8, RASA1
Degree	I	MAPK8, RHOA
	D	GRB2, PAK1, PRKCD, PRKCH, PRKCI, SORT1
Betweeness	I	None
	D	CR1, GRB2, RPS6KB1
Closeness	I	DLG4, SRC
	D	PRKCD, RAP1A, RPS6KA1

ClusCoef: clustering coefficient; I: increasing; D: decreasing.

1000 times) estimated to be nonzero and the average correlation between genes was 0.0023 (Additional file 1: Fig. S1).

Nodal centrality change during aD progression

There are totally 19 genes showing significant nodal centrality changes from CN to AD. For clustering coefficient, we found *C3*, *MAPK1*, *PAK1*, *PRKCH* and *SLA* with continuously increasing values from CN to EMCI, LMCI and AD. In contrast, *CALM3*, *MAPK8* and *RASA1* showed a decreasing pattern. For degree centrality, *MAPK8* and *RHOA* were found to increase and *GRB2*, *PAK1*, *PRKCD*, *PRKCH*, *PRKCI*, *SORT1* were found to decrease when subjects progress to a more severe stage. For betweenness, *CR1*, *GRB2* and *RPS6KB1* demonstrated a decreasing pattern. For closeness, *DLG4* and *SRC* were observed to

increase while *PRKCD*, *RAP1A* and *RPS6KA1* showed a decreasing pattern. Among these, early centrality changes are captured by *RPS6KB1*, *MAPK8*, *PRKCD*, *PRKCH* and *CR1* from CN to EMCI. A complete list of genes with continuous centrality changes is shown in Table 3. The detail centrality values of identified genes are in (Additional file 1: Tables S2–S5.) In Fig. 2 is the co-expression network module in the CN group that includes only genes showing gradual nodal changes. *RPS6KB1* is observed to be the top hub.

Co-expression change during aD progression

There are 66 pairs of genes that demonstrated a continuously decreasing or increasing co-expression from CN to EMCI, LMCI and AD. Based on their change patterns, we further divided these genes into eight clusters (Fig. 3). For clusters 1 and 2, correlation between those gene pairs shows continuous increase or decrease from CN to AD. For clusters 3 and 4, correlation between those gene pairs remains relatively stable until LMCI. On the contrary, gene pairs in clusters 5 and 7 show significant correlation change from CN to EMCI, but then remain stable until AD. After merging all of these modules, we observed several hub genes including *RPS6KB1* (Ribosomal Protein S6 Kinase B1), *SORT1*, *MAPK3*, textitPRKCD, *MAPK8* and *GRINA*. We further performed an pathway enrichment analysis for these clusters using hyper-geometric test (Additional file 1: Table S1). Five candidate pathways are significantly enriched for all clusters except cluster 7 and 8. Among eight gene clusters, early correlation
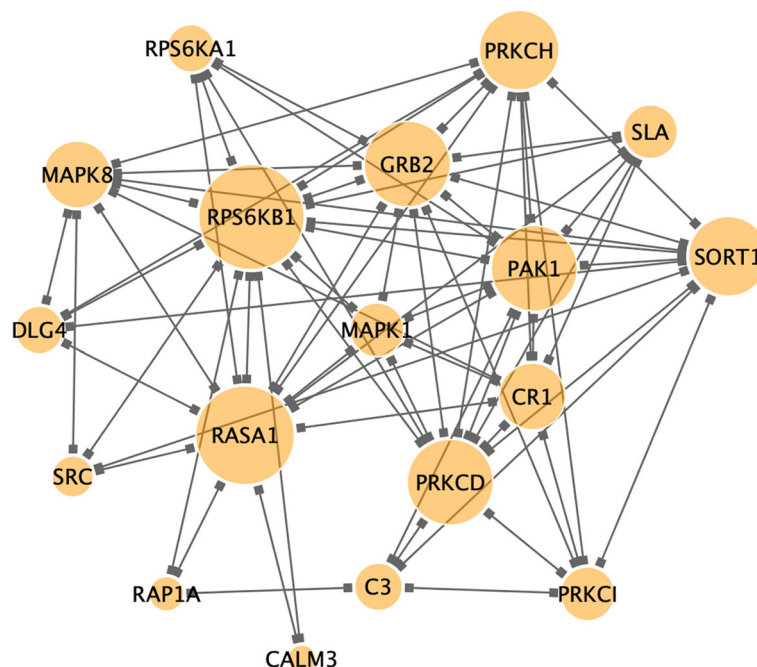
Fig. 2 Network module with genes showing continuous nodal centrality change from CN to AD

Upadhyaya *et al. BMC Medical Genomics* 2020, **13**(Suppl 5):53

Page 7 of 10

Fig. 3 Eight gene clusters with distinct patterns of gene co-expression change from CN to AD
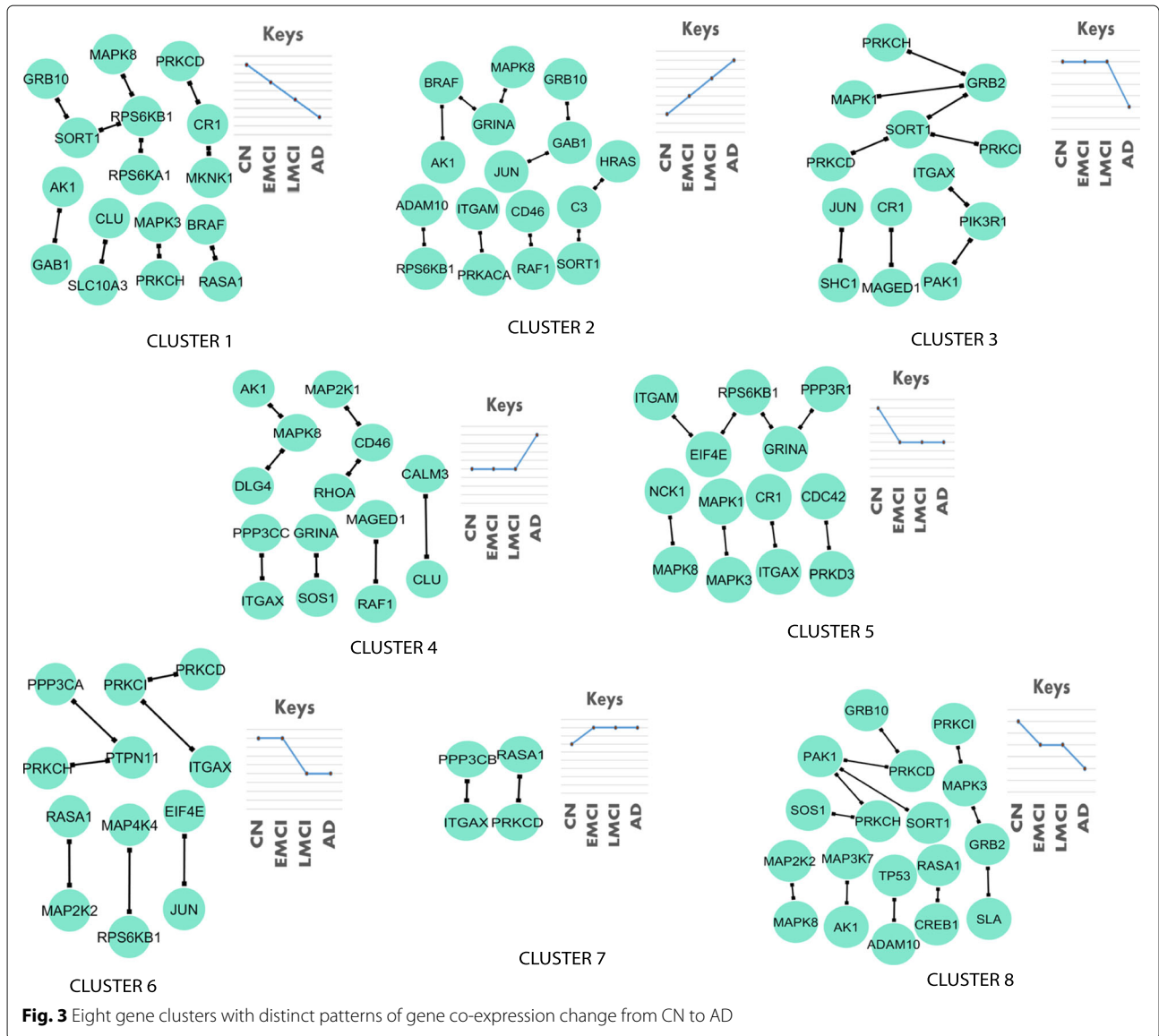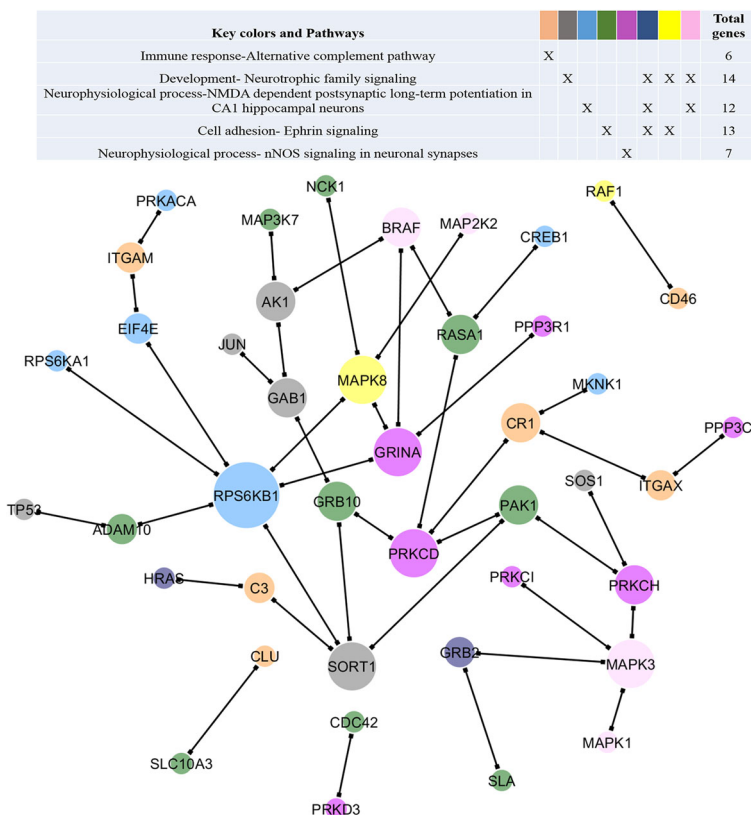
changes occurred in five of them, either decreasing or increasing. We combined these five gene clusters and formed a gene module as shown in Fig. 4. *RPS6KB1* (Ribosomal Protein S6 Kinase B1) was found to be the hub gene, followed by *GRINA*, *MAPK3*, *PRKCD*, *MAPK8* and *SORT1*.

Global network change during aD progression
Shown in Fig. 5 is the global clustering coefficient change of co-expression networks from CN to AD. It shows that the overall weighted clustering coefficient of the co-expression network remains relatively stable from CN to EMCI.

Discussion
The superior performance of modified JGL over traditional graphical lasso, when using the randomly permuted data, suggests that the similarity constraints on the co-expression networks of consecutive disease stages can help effectively control the probability to generate false positives. Therefore, the differential co-expression patterns identified through JGL results will provide more accurate information of the altered biological system during disease formation. On the other hand, while the high false positive rate of traditional graphical lasso could be controlled by the multiple test correction, it will be subject to the selection of correction method and significance

Upadhyaya *et al. BMC Medical Genomics* 2020, **13**(Suppl 5):53

Page 8 of 10

Key colors and Pathways									Total genes
Immune response-Alternative complement pathway	X								6
Development- Neurotrophic family signaling	X				X	X	X		14
Neurophysiological process-NMDA dependent postsynaptic long-term potentiation in CA1 hippocampal neurons			X			X		X	12
Cell adhesion- Ephrin signaling				X		X	X		13
Neurophysiological process- nNOS signaling in neuronal synapses					X				7

Fig. 4 Network module where gene co-expression (i.e.,links) exhibit early change from CN to EMCI. Node color indicates their membership in different pathways

threshold. In contrast, the similarity constraint used in JGL played a key role in controlling the false positive rate even without multiple correction. Therefore, JGL does not require as large number of permutations as traditional JGL and is more computationally cost-effective.

The global clustering coefficient change of co-expression networks from CN to AD suggests strong structural resilience of the co-expression network in the early development of AD. However, the global clustering coefficient drops significantly when patients progress
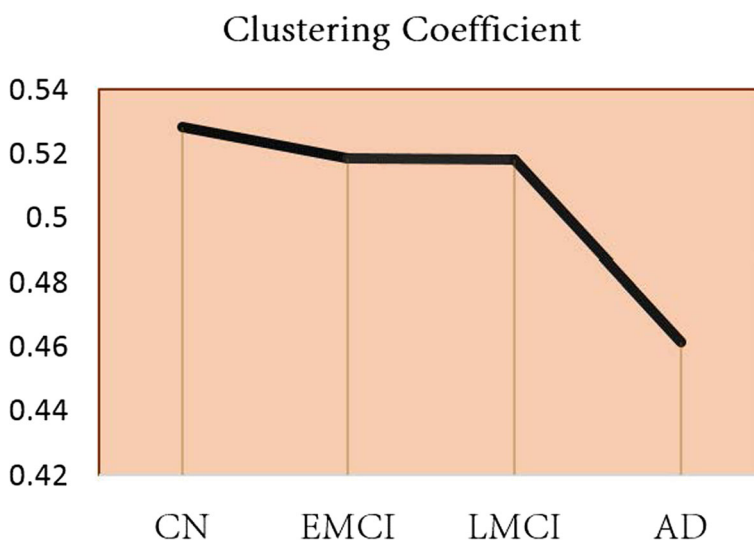
Fig. 5 Global weighted clustering co-efficient for co-expression networks derived from our JGL model

Upadhyaya *et al. BMC Medical Genomics* 2020, **13**(Suppl 5):53

Page 9 of 10

from LMCI to AD, which indicates the overall decoupling of genes in AD-enriched pathways starts dissolve at the late stage of AD. There is less probability to hold modular structures within co-expression network when the disease progresses from less severe to more severe stages.

Among all the genes showing significant changes in node or edge level metrics, *RPS6KB1* has consistently been observed as the top hub. Given that its centrality and co-expression changes occur in the early stage of AD, it holds great potential to serve as the biomarker for early diagnosis of AD. *RPS6KB1* encodes for a Ser/Thr (S/T)-directed kinase, 70-kDa S6 kinase (p70S6K), that plays a crucial role in cell growth, cell differentiation, and cell cycle control. The p70S6K can phosphorylate tau at S262, S214, and T212 sites [22]. It is found to be highly expressed in cerebral cortex and hippocampus, a brain region affected by AD, according to the Human Protein Atlas (HPA) [23]. In a large-scale genome wide association study (GWAS), it was reported to be a genetic risk loci for multiple sclerosis [24]. Findings from earlier studies suggest that the additive effect of alleles in *RPS6KB1* and several other genes in tau kinase pathway are associated with late-onset of AD in *APOE* non-e4 carriers [22, 25].

MAPK8, also known as *JNK1*, is one of the three proteins in c-Jun N-terminal Kinase family. It is identified to be involved in the regulation and maintenance of physiological responses in central nervous system [26]. Studies have also found positive correlation between JNK phosphorylation and expression of amyloid beta (Aβ) peptide, which is a well-known AD hallmark[27].

Other top hub genes that also show early changes in nodal centrality or co-expression patterns are *MAPK8,MAPK3, GRINA, PRKCD, SORT1, RASA1* and *PAK1*. A few studies have reported the association of these genes with AD. For example, [28] suggests a potential causative role of P21 activated kinases defects on the cognitive deficits of AD patients. However, their coordination has not been previously studied yet. This work is the first study to reveal how their co-expression pattern changes during AD progression. Given that the co-expression and nodal centrality of these genes start to change in the CN stage, they have great potential to serve as systems biomarker to capture the biological alterations in the very early stage of AD and further help with the development of AD therapeutic intervention.

Conclusion

We employed a modified joint graphical lasso for estimation of co-expression networks for multiple stages of AD. Comparing with graphical lasso, JGL accounts for the similarity in consecutive disease stages. Our results on random data sets shows that it is less likely to generate false positives. In the subsequent differential co-expression analysis, we found that the clustering coefficient of the co-expression network shows significant changes only when subjects progress from LMCI to AD. Node wise and edge wise comparison have led to eight gene clusters that demonstrate continuous changes from CN, EMCI to LMCI and AD. Particularly, five of them shows differential co-expression patterns from CN to EMCI. Genes in these modules could be used as systems biomarkers for early screening in AD. However, more efforts are warranted to validate their downstream function.

Supplementary information

> **Additional file 1:** Differential co-expression analysis reveals early stage transcriptomic decoupling in Alzheimer's disease.

Upadhyaya *et al. BMC Medical Genomics* 2020, **13**(Suppl 5):53

Page 10 of 10

Availability of data and materials
Demographic information, gene expression data, and diagnostic information are available from the ADNI data repository at http://adni.loni.usc.edu/data-samples/access-data/.

Ethics approval and consent to participate
This work does not involve any new data collection. It focuses on the analysis of existing data sets in the ADNI. More information about the ethics approval for the ADNI data can be found in https://adni.loni.usc.edu/wp-content/uploads/2017/09/ADNID_Approved_Protocol_11.19.14.pdf.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of BioHealth Informatics, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA. [2]Department of Electric and Computer Engineering, Indiana University Purdue University Indianapolis Indianapolis, IN, USA. [3]Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, USA. [4]Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA.

References
1. Association A, et al. 2018 alzheimer's disease facts and figures. Alzheim Demen. 2018;14(3):367–429.
2. Liu C-C, Kanekiyo T, Xu H, Bu G. Apolipoprotein e and alzheimer disease: risk, mechanisms and therapy. Nat Rev Neurol. 2013;9(2):106.
3. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, DeStefano AL, Bis JC, Beecham GW, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for alzheimer's disease. Nat Genet. 2013;45(12):1452.
4. Kang H, Lee J, Yu S. Differential co-expression networks using rna-seq and microarrays in alzheimer's disease. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2016. https://doi.org/10.1109/bibm.2016.7822811.
5. Wang Z, Yan X, Zhao C. Dynamical differential networks and modules inferring disrupted genes associated with the progression of alzheimer's disease. Exp Ther Med. 2017;14(4):2969–75.
6. Langfelder P, Horvath S. Wgcna: an r package for weighted correlation network analysis. BMC bioinformatics. 2008;9(1):559.
7. Lauritzen SL. Graphical Models vol. 17. Oxford: Clarendon Press; 1996.
8. Danaher P, Wang P, Witten DM. The joint graphical lasso for inverse covariance estimation across multiple classes. J R Stat Soc Ser B Stat Methodol. 2014;76(2):373–97.
9. Bu X, Xiang Y, Jin W, Wang J, Shen L, Huang Z, Zhang K, Liu Y, Zeng F, Liu J, et al. Blood-derived amyloid-β protein induces alzheimer's disease pathologies. Mol Psychiatry. 2018;23(9):1.
10. Lunnon K, Keohane A, Pidsley R, Newhouse S, Riddoch-Contreras J, Thubron EB, Devall M, Soininen H, Kłoszewska I, Mecocci P, et al. Mitochondrial genes are altered in blood early in alzheimer's disease. Neurobiol Aging. 2017;53:36–47.
11. Blennow K. A review of fluid biomarkers for alzheimer's disease: moving from csf to blood. Neurol Ther. 2017;6(1):15–24.
12. Proitsi P, Kim M, Whiley L, Simmons A, Sattlecker M, Velayudhan L, Lupton MK, Soininen H, Kloszewska I, Mecocci P, et al. Association of blood lipids with alzheimer's disease: A comprehensive lipidomics analysis. Alzheim Demen. 2017;13(2):140–51.
13. Shen L, Thompson PM, Potkin SG, Bertram L, Farrer LA, Foroud TM, Green RC, Hu X, Huentelman MJ, Kim S, et al. Genetic analysis of quantitative phenotypes in ad and mci: imaging, cognition and biomarkers. Brain Imaging Behav. 2014;8(2):183–207.
14. Xulvi-Brunet R, Li H. Co-expression networks: graph properties and topological comparisons. Bioinformatics. 2009;26(2):205–14.
15. Azuaje FJ. Selecting biologically informative genes in co-expression networks with a centrality score. Biol Direct. 2014;9(1):12.
16. Saykin AJ, Shen L, Yao X, Kim S, Nho K, Risacher SL, Ramanan VK, Foroud TM, Faber KM, Sarwar N, et al. Genetic studies of quantitative mci and ad phenotypes in adni: Progress, opportunities, and plans. Alzheim Demen. 2015;11(7):792–814.
17. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends® Mach Learn. 2011;3(1):1–122.
18. Efron B, Hastie T, Johnstone I, Tibshirani R, et al. Least angle regression. Ann Stat. 2004;32(2):407–99.
19. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proc Nat Acad Sci. 2004;101(11):3747–52.
20. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. Soc Netw. 2010;32(3):245–51.
21. Opsahl T. tnet: Software for analysis of weighted, two-mode, and longitudinal networks. R package. 2007.
22. Pei J-J, Björkdahl C, Zhang H, Zhou X, Winblad B. p70 s6 kinase and tau in alzheimer's disease. J Alzheim Dis. 2008;14(4):385–92.
23. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, et al. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
24. International Multiple Sclerosis Genetics Consortium. Manba, cxcr5, sox8, rps6kb1 and zbtb46 are genetic risk loci for multiple sclerosis. Brain. 2013;136(6):1778–82.
25. Vázquez-Higuera JL, Mateo I, Sánchez-Juan P, Rodríguez-Rodríguez E, Pozueta A, Calero M, Dobato JL, Frank-García A, Valdivieso F, Berciano J, et al. Genetic variation in the tau kinases pathway may modify the risk and age at onset of alzheimer's disease. J Alzheim Dis. 2011;27(2):291–7.
26. Yarza R, Vela S, Solas M, Ramirez MJ. c-jun n-terminal kinase (jnk) signaling as a therapeutic target for alzheimer's disease. Front Pharmacol. 2016;6:321.
27. Ma Q-L, Yang F, Rosario ER, Ubeda OJ, Beech W, Gant DJ, Chen PP, Hudspeth B, Chen C, Zhao Y, et al. β-amyloid oligomers induce phosphorylation of tau and inactivation of insulin receptor substrate via c-jun n-terminal kinase signaling: suppression by omega-3 fatty acids and curcumin. J Neurosci. 2009;29(28):9078–89.
28. Zhao L, Ma Q-L, Calon F, Harris-White ME, Yang F, Lim GP, Morihara T, Ubeda OJ, Ambegaokar S, Hansen JE, et al. Role of p21-activated kinase pathway defects in the cognitive deficits of alzheimer disease. Nat Neurosci. 2006;9(2):234.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.