Featured Article

Systems biology approach to late-onset Alzheimer's disease genome-wide association study identifies novel candidate genes validated using brain expression data and *Caenorhabditis elegans* experiments

Shubhabrata Mukherjee[a,*], Joshua C. Russell[b], Daniel T. Carr[b], Jeremy D. Burgess[c], Mariet Allen[c], Daniel J. Serie[d], Kevin L. Boehme[e,f], John S. K. Kauwe[e,f], Adam C. Naj[g], David W. Fardo[h], Dennis W. Dickson[c], Thomas J. Montine[b,§], Nilufer Ertekin-Taner[c,i], Matt R. Kaeberlein[b], Paul K. Crane[a]

[a]*Department of Medicine, University of Washington, Seattle, Washington, USA*
[b]*Department of Pathology, University of Washington, Seattle, Washington, USA*
[c]*Department of Neuroscience, Mayo Clinic Florida, Jacksonville, Florida, USA*
[d]*Department of Health Sciences Research, Mayo Clinic Florida, Jacksonville, Florida, USA*
[e]*Department of Biology, Brigham Young University, Provo, Utah, USA*
[f]*Department of Neuroscience, Brigham Young University, Provo, Utah, USA*
[g]*Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA*
[h]*Department of Biostatistics, University of Kentucky, Lexington, Kentucky, USA*
[i]*Department of Neurology, Mayo Clinic Florida, Jacksonville, Florida, USA*

Abstract

Introduction: We sought to determine whether a systems biology approach may identify novel late-onset Alzheimer's disease (LOAD) loci.

Methods: We performed gene-wide association analyses and integrated results with human protein-protein interaction data using network analyses. We performed functional validation on novel genes using a transgenic *Caenorhabditis elegans* Aβ proteotoxicity model and evaluated novel genes using brain expression data from people with LOAD and other neurodegenerative conditions.

Results: We identified 13 novel candidate LOAD genes outside chromosome 19. Of those, RNA interference knockdowns of the *C. elegans* orthologs of *UBC, NDUFS3, EGR1,* and *ATP5H* were associated with Aβ toxicity, and *NDUFS3, SLC25A11, ATP5H,* and *APP* were differentially expressed in the temporal cortex.

Discussion: Network analyses identified novel LOAD candidate genes. We demonstrated a functional role for four of these in a *C. elegans* model and found enrichment of differentially expressed genes in the temporal cortex.

© 2017 The Alzheimer's Association. Published by Elsevier Inc. All rights reserved.

Keywords: Alzheimer's disease; SNP; Protein-protein interaction; *C. elegans*; Brain expression; Network analysis; Systems biology

---

1. Introduction

Most late-onset Alzheimer's disease (LOAD) genetic research has pursued one variant at a time approaches such as genome-wide association studies (GWASs). Lambert et al. [1] published the largest LOAD GWAS to date and identified about two dozen loci associated with LOAD.

Although GWAS is an important first step, additional approaches will also likely contribute to understanding the genetic determinants of LOAD. A three-component approach [2]—GWAS, gene, and network/pathway-based analyses—has been recommended to more fully characterize genetic architecture of complex diseases.

Previously, network analyses using gene expression data from 1647 postmortem brain tissues from LOAD patients and nondemented individuals have found an immune and microglia-specific module [3]. Immune response, regulation of endocytosis, cholesterol transport, and protein ubiquitination pathways were significant [4].

The strategy to integrate human protein-protein interaction (PPI) data with gene-wide association results strategy implemented here refines the gene-based approach by incorporating additional biological knowledge. This approach capitalizes on the idea that protein-encoding genes known to interact with multiple other proteins tend to be associated with more extensive regulation and are more likely to cause complex pathologic processes than genes with fewer interactions [5,6].

In this article, we use a dense module search (DMS) approach using human PPI data to prioritize gene-based analyses of GWAS results. We evaluate the plausibility of the resulting network module using experiments with transgenic *Caenorhabditis elegans* models of β-amyloid (Aβ) aging-related proteotoxicity and brain expression data.

2. Methods

We present a flowchart of the analytic steps in Supplementary Fig. 1.

2.1. Stage 1 data: GWAS results

The LOAD GWAS data set was reported by the International Genomics of Alzheimer's Project (IGAP) Consortium [1]. These data were derived from 17,008 people with LOAD and 37,154 cognitively normal elderly control subjects. IGAP includes data from the Alzheimer's Disease Genetics Consortium, the Genetic and Environmental Risk in Alzheimer's Disease Consortium, the European Alzheimer's Disease Initiative, and the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (see [1] for more details). Full details are provided in Supplementary Section 1.

The IGAP analysis included single-nucleotide polymorphisms (SNPs) with minor allele frequencies ≥0.01 and the imputation quality score ≥0.3 in each study, resulting in 7,055,881 SNPs. The SNPs allelic association result file is available from http://www.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php.

2.2. Gene-wide analysis

We used the Versatile Gene-Based Test for Genome-wide Association Study [7] (VEGAS) routine in Fast Association

Tests [8] for gene-wide analysis. See Supplementary Section 3 for further details on VEGAS. We used all SNPs within ±50 kb of the untranslated regions (UTRs) for each gene. We used NCBI Build 37 to assign SNPs to 34,211 genes and pseudogenes (hereinafter "genes"). We retained 6,753,292 of the 7,055,811 SNPs that passed QC (95.7%). These SNPs were mapped to 33,086 genes with 1–15,373 SNPs per gene.

We repeated gene-wide association analyses using a more stringent ±0 kb of the UTRs as a sensitivity analysis, resulting in 28,370 genes.

2.3. DMS-based analyses

We mined human interactome PPI data (190,526 unique interactions for 15,260 genes based on biological evidence) using the R package *iRefR* [9]. The *iRefR* provides an index of protein interactions available in primary interaction databases: BIND, BioGRID, CORUM, DIP, HPRD, InnateDB, IntAct, MatrixDB, MINT, MPact, MPIDB, MPPI, and OPHID (all acronyms, citations and URLs are in Supplementary Section 2). There were 13,550 genes in common between the gene-wide analysis and the PPI databases. Not all genes code for proteins and of those that do, not all have interactions with proteins of other genes. Many genes do not have any IGAP SNPs that map within the ±50 kb boundary.

We integrated gene-wide results with the PPI data using dense module GWAS (dmGWAS) [10] to identify candidate genes and subnetworks. dmGWAS, a DMS method, identifies networks of interacting genes enriched with low *P* values by searching the entire interactome and exhaustively examining the combined effect of multiple genes. See Supplementary Section 4 for details of the DMS method used by dmGWAS.

We used Cytoscape [11] to visualize the top module as an undirected graph using the "betweenness centrality" measure, defined as the length of shortest paths from all nodes to all other nodes. We report the betweenness centrality measure and the degree of each gene/node in Table 1. The degree of a node/gene represents the number of edges (connecting two nodes/genes) linked to that node/gene.

As a sensitivity analysis, we performed the DMS analysis omitting the gene with the highest degree and betweenness centrality measure.

2.4. Evaluation of novel candidate LOAD genes with a transgenic C. elegans model

We assessed the roles of novel candidate genes in the top module outside chromosome 19 that had nematode orthologs. We used transgenic *C. elegans* models in which either the Aβ$_{3-42}$ (CL2006) or the Aβ$_{1-42}$ (GMC101) (both referred to subsequently as "Aβ") peptide was expressed in body wall muscle cells under the control of the *unc-54* promoter [13,14]. Culture of transgenic *C. elegans*, RNA interference

Table 1
Gene-wide and DMS results for the genes identified in the top DMS network module

Chr	Gene	Results from VEGAS		Results from DMS network analysis	
		SNPs	P value	Betweenness centrality	Degree
2	*BIN1*	607	$<1.0 \times 10^{-6}$	0.003	2
4	*ALB*	263	.40	0.129	6
5	*EGR1*	216	7.1×10^{-3}	0.124	4
6	*HLA-DRA*	1686	$<1.0 \times 10^{-6}$	0	2
6	*HLA-DRB1*	3388	$<1.0 \times 10^{-6}$	0	2
8	*CHRNA2*	341	7.2×10^{-6}	0	1
8	*MYC*	228	.02	0.025	5
11	*NDUFS3*	100	2.0×10^{-6}	0.001	3
11	*MS4A2*	249	$<1.0 \times 10^{-6}$	0	1
11	*PICALM*	556	$<1.0 \times 10^{-6}$	0	1
12	*UBC*	367	.35	0.728	23
17	*SLC25A11*	250	2.0×10^{-4}	0.063	4
17	*C1QBP*	366	4.1×10^{-3}	0.006	5
17	*KRT14*	191	.03	0.073	3
17	*ICT1*	233	8.8×10^{-4}	0.037	7
17	*ATP5H*	292	4.8×10^{-4}	0.097	4
19	*PVR*	302	$<1.0 \times 10^{-6}$	0	1
19	*BCL3*	272	$<1.0 \times 10^{-6}$	0	2
19	*CBLC*	301	$<1.0 \times 10^{-6}$	0	1
19	*BCAM*	301	$<1.0 \times 10^{-6}$	0	1
19	*PVRL2*	385	$<1.0 \times 10^{-6}$	0	1
19	*TOMM40*	318	$<1.0 \times 10^{-6}$	0.002	3
19	*APOE*	295	$<1.0 \times 10^{-6}$	0	3
19	*APOC1*	273	$<1.0 \times 10^{-6}$	0	1
19	*APOC2*	246	$<1.0 \times 10^{-6}$	0	1
19	*CLPTM1*	276	$<1.0 \times 10^{-6}$	0	2
19	*RELB*	270	$<1.0 \times 10^{-6}$	0.001	3
19	*CLASRP*	251	$<1.0 \times 10^{-6}$	0	1
19	*GEMIN7*	214	$<1.0 \times 10^{-6}$	0	1
19	*MARK4*	403	$<1.0 \times 10^{-6}$	0	1
19	*FBXO46*	200	$<1.0 \times 10^{-6}$	0	1
19	*FOXA3*	355	8.5×10^{-4}	0	1
21	*APP*	997	.03	0.175	9

Abbreviations: DMS, dense module search; Chr, chromosome; Gene, gene symbol; SNPs, number of single-nucleotide polymorphisms in the gene; *P* value, gene-wide *P* value from Versatile Gene-Based Test for Genome-wide Association Study (VEGAS); Betweenness centrality, a measure denoting the number of shortest paths from all vertices/interactions to all others that pass through that node/gene; Degree, the number of edges linked to that node/gene.

NOTE. In Fast Association Tests, the simulations for the VEGAS subroutine stop once a gene is genome-wide significant. In a previous article [12], VEGAS's authors established 1.0×10^{-6} as the significance threshold for VEGAS (corresponding to 0.05 genome-wide false-discovery rate).

(RNAi) knockdown, and assessment of age-associated paralysis were performed as previously described [14,15]. Worms were scored as paralyzed if they were unable to make forward progress on the surface of the nematode growth medium in response to plate tapping or tail prodding. All RNAi clones were verified by sequencing. RNAi was initiated from the fourth larval stage (L4). We assessed functional roles of novel genes using corresponding nematode orthologs and determined the effect of RNAi knockdown on toxicity caused by transgenic expression of the Aβ peptide. Statistical significance (α = 0.01) was determined using a Wilcoxon rank sum test.

2.5. Evaluation of novel candidate LOAD genes with human brain gene expression data

We also evaluated novel candidate genes in the top module outside chromosome 19 using brain gene expression data [16]. Gene expression data were available from the temporal cortex of 399 individuals and cerebellum of 374 individuals [17]. Complete methods are described in [16]. Briefly, RNA was isolated and its quantity and quality were determined [18]. Transcript levels were measured using Illumina whole-genome cDNA-mediated annealing, selection, and ligation (DASL) assays. Normalized differential expression levels were assessed for LOAD versus all non-LOAD individuals and versus those with progressive supranuclear palsy (PSP).

We used linear regression models with LOAD versus non-LOAD or versus PSP as the predictor and expression levels as endophenotypes, adjusting for the number of *APOE* ε4 alleles, age at death, sex, plate, RNA integrity number, and adjusted RNA integrity number squared. Further details are provided in Supplementary Section 5. Results are reported as false-discovery rate *q* values [19] following correction for the number of genes (and probes) evaluated.

3. Results

3.1. Variants identified by VEGAS and DMS approaches

VEGAS gene-wide results were similar to previously published GWAS results [1] (Supplementary Table 1). In addition to LOAD genes identified by prior GWAS and two additional genes identified in a previous analysis of IGAP data [20], VEGAS analyses identified novel signals for three genes and four pseudogenes, all *P* value $<1.0 \times 10^{-6}$ except as indicated: the genes *HBEGF* (chromosome 5; *P* value = 2.0×10^{-6}), *SLC4A9* (chromosome 5), and *HLA-DRA* (chromosome 6), and the pseudogenes *CDCA4P3* (chromosome 1), *GULOP* (chromosome 6), *YWHAZP9* (chromosome 11), and *SLC25A1P1* (chromosome 11) (see Supplementary Table 1).

P values for most genes in our sensitivity analysis (±0 kb of the UTR map) were similar (see Supplementary Table 1) compared with the ±50 kb of the UTR mapping scheme.

The top DMS module contained 33 unique genes with 53 interactions (see Fig. 1). Many of these genes were on chromosome 19 and may represent linkage disequilibrium with *APOE*. Seventeen were not on chromosome 19, of which four were previously identified to be associated with LOAD risk (*BIN1*, *HLA-DRB1*, *MS4A2*, and *PICALM*) and 13 have not been previously identified in the GWAS [1] or prior gene-based analyses [20] of LOAD: *ALB*, *EGR1*, *HLA-DRA*, *CHRNA2*, *MYC*, *NDUFS3*, *UBC*, *SLC25A11*, *C1QBP*, *KRT14*, *ICT1*, *ATP5H*, and *APP* (see Table 1).
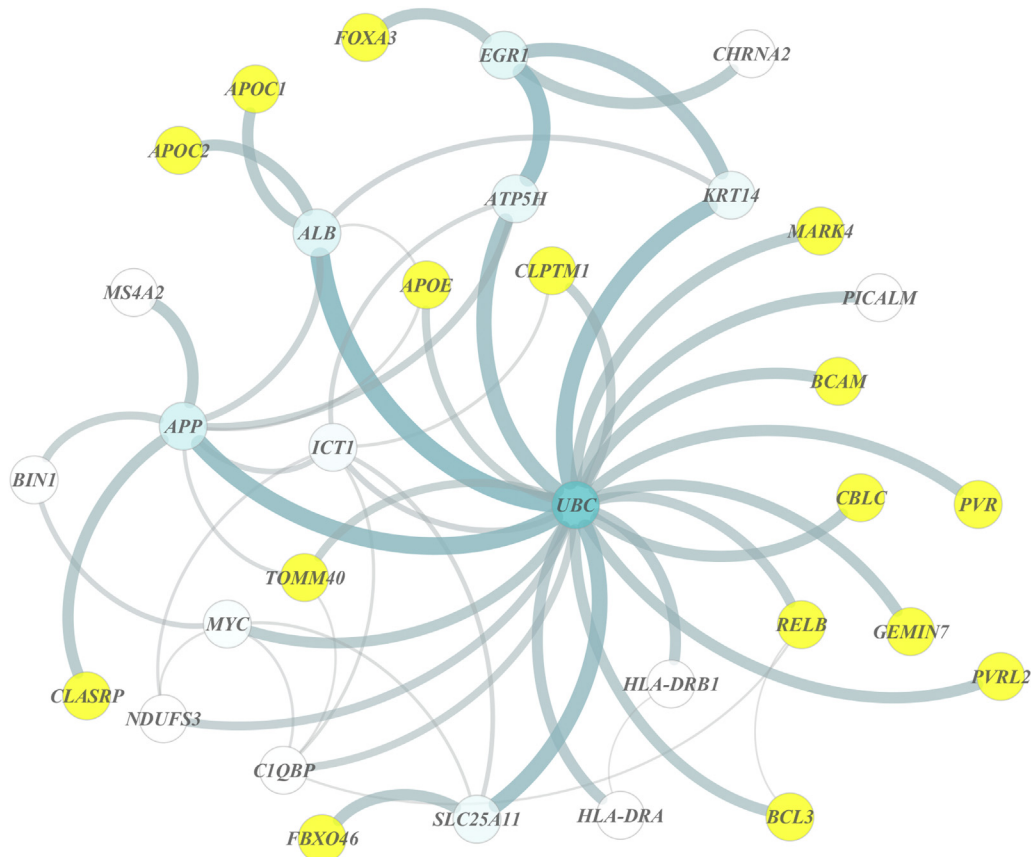
Fig. 1. Protein-protein interaction (PPI) subnetwork visualization for the top dense module searching (DMS)-identified module. In DMS, PPI data are used to identify groups of genes ("modules") whose products interact with each other. These groups of genes are compared on the basis of having a higher proportion with statistically significant associations with a phenotype, which in this case is late-onset Alzheimer's disease. The darkness of the nodes and the thickness of the vertices of the nodes both reflect the "betweenness centrality" measure, which is defined as the number of shortest paths from all vertices/interactions to all others that pass through that node/gene. The "degree" of each gene is represented by the number of vertices connecting it with other genes. This figure suggests that *UBC* is highly relevant in this network, and that it may act as an important organizing regulatory molecule in this biological network. Genes in chromosome 19 are highlighted in yellow. The color version of this figure is available in the online edition.

Network analysis graphs for the top three and top five modules are shown in Supplementary Figs. 2 and 3. The top three and five modules included 49 and 74 unique genes.

Results from our ±0 kb sensitivity analyses are presented in Supplementary Table 2.

UBC, APP, and *ALB* had the highest betweenness centrality (0.728, 0.175, and 0.129) and degree (23, 9, and 6) values. *UBC* had much higher values than any other gene, so we were concerned that it could be driving our results. Sensitivity analyses excluding *UBC* resulted in a top module with 47 unique genes and 71 interactions anchored by *APP* and *MYC* (see Supplementary Fig. 4). Twenty-six of the 33 genes in the top module from our primary analyses were also present in the *UBC*-free sensitivity analyses. The six genes whose status depended on *UBC* were *ATP5H, EGR1, KRT14, CHRNA2, C1QBP*, and *FOXA3*.

3.2. C. elegans *results*

We identified *C. elegans* orthologs for four of the 13 novel genes in the top module: *UBC, ATP5H, EGR1*, and *NDUFS3*. In addition, we identified orthologs of two well-known LOAD loci: *BIN1* and *PICALM*.

RNAi knockdown of *C. elegans UBC* orthologs (*ubq-1* and *ubq-2* are targeted by a single RNAi clone) significantly accelerated age-associated onset of Aβ_{3-42} toxicity (see Fig. 2A; $P < .01$). RNAi knockdown of *NDUFS3* (*nuo-2*) and *ATP5H* (*atp-5*) *C. elegans* orthologs significantly delayed paralysis because of Aβ_{3-42} toxicity (see Fig. 2B and C; both P values <.01). RNAi knockdowns of *EGR1* (*egrh-1*), *BIN1* (*amph-11*), and *PICALM* (*unc-11*) *C. elegans* orthologs significantly delayed paralysis because of Aβ_{1-42} toxicity (see Fig. 2E; all P values <.001). None of the RNAi conditions induced paralysis in the control worms (CL2122) for the aforementioned six genes (see Fig. 2D and F).

3.3. *Brain expression results*

Data were available for probes that targeted 11 of the 13 novel genes outside chromosome 19; *CHRNA2* and *KRT14* had low expression levels. The 11 genes were targeted by 15 probes. Four (*NDUFS3, SLC25A11, ATP5H*, and *APP*) of the 11 genes (36%) had differentially expressed probes
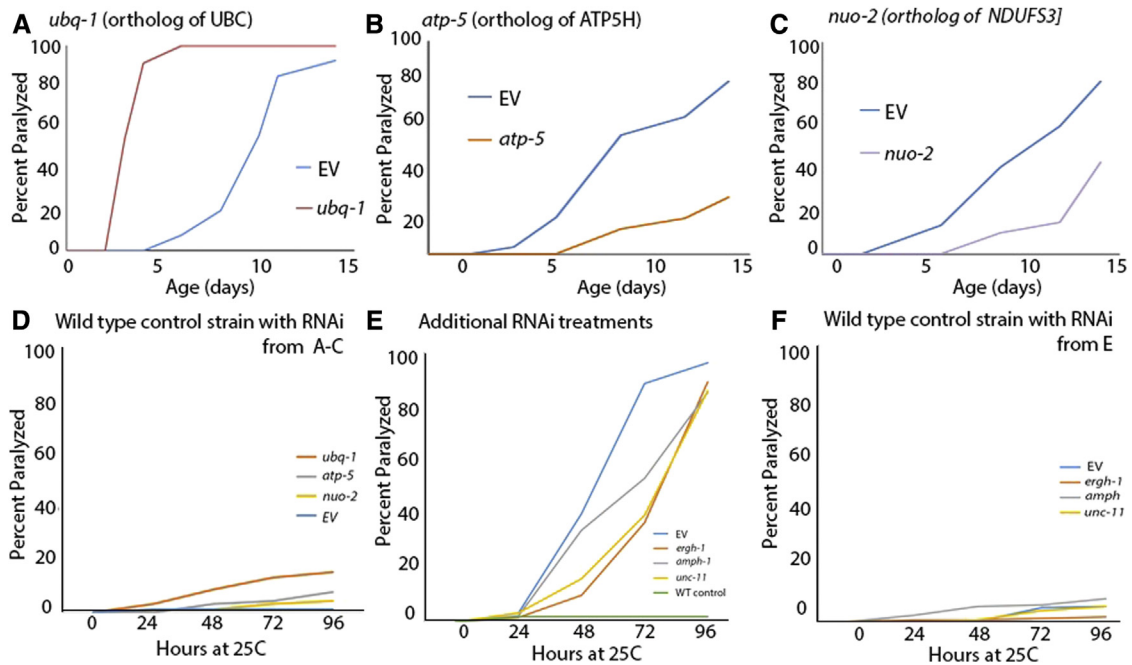
Fig. 2. Graphs of the proportion of Aβ-expressing transgenic *Caenorhabditis elegans* strains CL2006 and GMC101 exhibiting changes in age-related paralysis with RNAi knockdown experiments. RNAi knockdowns of *ubq-1* (A) exacerbate paralysis, whereas RNAi knockdowns of *atp-5* and *nuo-2* (B, C) reduce paralysis in Aβ$_{3-42}$-expressing *C. elegans* strain CL2006. Wild type (WT) control strain CL2122 (D) does not become paralyzed from RNAi knockdowns of genes evaluated in A to C. RNAi knockdowns of *ergh-1, amph-1,* and *unc-11* reduce paralysis (E) in Aβ$_{1-42}$-expressing *C. elegans* strain GMC101. RNAi knockdowns of genes evaluated in E do not result in paralysis (F) of WT control strain CL2122. RNAi or empty vector (EV) was initiated at the fourth larval stage (L4) just before adulthood. The data plotted in these graphs indicate the proportion of worms found to be paralyzed at each time point.

($q < 0.05$) in the temporal cortex. This figure is enriched compared with all expressed probes in the same experiment: 1933 of 13,592 (14%) probes had differential expression. Two additional genes (*UBC* and *C1Q8P*) had differential expression in cerebellum (Table 2).

We also compared gene expression for people with LOAD to the group with PSP; six of the 11 genes (55%) had differential expression in the cortex ($q < 0.05$), including the four genes with differential cortical expression and the two genes with differential cerebellar expression in comparison with all non-AD neurodegeneration (Table 2).

SLC25A11 expression levels were the most different (Table 2). Cortical expression levels were lower in people with LOAD than for people with other neurodegenerative conditions for all differentially expressed genes except *APP*.

4. Discussion

We identified three novel genes using gene-wide analyses. The novel locus *HBEGF* (chromosome 5: heparin-binding epidermal growth factor-like growth factor) is recognized as an important component for the modulation of cell activity. Found widely distributed in cerebral neurons and neuroglia, *HBEGF* induced by brain hypoxia and/or ischemia subsequently stimulates neurogenesis [21]. The protein encoded by *SLC4A9* (chromosome 5; solute carrier family 4, sodium bicarbonate

cotransporter, member 9), a neighbor of *HBEGF*, is a membrane protein involved in anion exchange expressed primarily in kidney [22]. *HLA-DRA* major histocompatibility complex, class II, DR alpha (chromosome 6) is an HLA class II alpha chain paralogue. *HLA* associations have been previously reported in Alzheimer's disease [1], Parkinson's disease [23,24], and multiple sclerosis [25,26]. In a recent article, the *HLA* locus provided support to the notion of a link between frontotemporal dementia and the immune system. Analyses of DNA methylation data suggested risk at that locus was associated with *cis*-changes in methylation levels of *HLA-DRA* in frontal cortex [27].

The DMS analysis identified genes previously associated with LOAD and genes not previously associated with LOAD. Five genes—*UBC, APP, EGR1, ALB,* and *ATP5H*—were highly relevant in the top module as indicated by high betweenness centrality measure values. Four genes that had not previously been associated with LOAD had nematode orthologs, and we completed experiments on four of these. We were able to validate the associations of *UBC, ATP5H, EGR1,* and *NDUFS3* using a *C. elegans* model of age-related Aβ toxicity. A high proportion of genes not previously associated with LOAD had differential expression in the temporal cortex from people with LOAD compared with people with non-LOAD neurodegeneration because of PSP.

Table 2
Differences in the expression in brains of individuals with (A) LOAD pathology versus all non-LOAD individuals and (B) LOAD pathology versus those with PSP for overlapping genes between DMS top module and brain expression results

| | | | LOAD versus non-LOAD | | | | | | LOAD versus PSP | | | | | |
| | | | Temporal cortex | | | Cerebellum | | | Temporal cortex | | | Cerebellum | | |
Chr	Probe	Gene symbol	β	SE	q value	β	SE	q value	β	SE	q value	β	SE	q value
4	ILMN_1782939	*ALB*	−0.07	0.08	0.52	−0.17	0.05	0.02	0.02	0.09	0.89	−0.11	0.06	0.43
5	ILMN_1762899	*EGR1*	−0.05	0.04	0.37	−0.001	0.06	0.99	−0.01	0.05	0.89	0.08	0.07	0.77
6	ILMN_2157441	*HLA-DRA*	−0.01	0.07	0.98	0.002	0.09	0.99	0.06	0.09	0.62	0.06	0.10	0.86
8	ILMN_2110908	*MYC*	0.11	0.05	0.12	0.07	0.06	0.59	0.06	0.07	0.50	0.02	0.07	0.90
11	ILMN_1756355	*NDUFS3*	−0.03	0.01	2.1×10^{-3}	−0.003	0.01	0.96	−0.05	0.01	2.9×10^{-5}	−0.01	0.01	0.86
12	ILMN_2331501	*UBC*	−0.03	0.01	0.06	−0.01	0.01	0.96	−0.05	0.01	0.01	−0.01	0.01	0.82
12	ILMN_2252160	*UBC*	−0.01	0.08	0.98	0.04	0.06	0.88	−0.17	0.10	0.15	0.03	0.07	0.86
17	ILMN_1664168	*SLC25A11*	−0.08	0.02	9.9×10^{-5}	−0.02	0.01	0.49	−0.12	0.02	9.3×10^{-7}	−0.01	0.02	0.86
17	ILMN_1668996	*C1QBP*	−0.04	0.01	0.05	−0.01	0.02	0.96	−0.06	0.02	3.1×10^{-3}	0.01	0.02	0.90
17	ILMN_2182198	*ICT1*	−0.04	0.02	0.19	0.01	0.02	0.88	−0.06	0.03	0.06	0.02	0.02	0.80
17	ILMN_1666372	*ATP5H*	−0.04	0.01	0.01	−0.01	0.01	0.75	−0.08	0.02	5.2×10^{-5}	−0.02	0.02	0.72
17	ILMN_1794912	*ATP5H*	−0.05	0.02	0.02	0.02	0.02	0.50	−0.06	0.02	0.01	0.01	0.02	0.83
21	ILMN_2404063	*APP*	0.07	0.03	0.07	0.02	0.03	0.88	0.09	0.04	0.03	0.01	0.03	0.90
21	ILMN_2404065	*APP*	0.05	0.02	0.03	−0.01	0.02	0.96	0.05	0.02	0.04	0.01	0.02	0.88
21	ILMN_1653283	*APP*	0.12	0.04	0.02	−0.004	0.03	0.98	0.09	0.04	0.05	−0.02	0.04	0.86

Abbreviations: DMS, dense module search; Chr, chromosome; LOAD, late-onset Alzheimer's disease; PSP, progressive supranuclear palsy.
NOTE. Beta coefficients and *q* values of association refer to difference in expression levels between (A) LOAD and people without LOAD, and (B) LOAD and people with PSP.

UBC codes for ubiquitin C, a polyubiquitin precursor. Ubiquitinization is an important process that promotes synaptic integrity [28], thought to be of critical importance in LOAD pathobiology, and a feature of the characteristic neuropathologic features of LOAD. The *UBC* finding here is specific to *UBC* and not to other ubiquitin pathway genes. The *C. elegans* model for Aβ$_{3-42}$ toxicity was sensitive to *UBC* knockdown with RNAi, such that knockdown of *UBC* orthologs significantly accelerated the age-associated onset of Aβ$_{3-42}$ toxicity. *UBC* expression levels in the temporal cortex were lower in individuals with LOAD than people with non-LOAD neurodegeneration. Taken together, these data strongly suggest that *UBC* may be an important locus in the genetic architecture of LOAD, where reduced levels of *UBC* may lead to LOAD risk via a mechanism that enhances Aβ toxicity in vulnerable brain regions.

Although *UBC* has many interactions with other genes in the top module, it did not appear to be necessary for our results, as the top module excluding *UBC* in our sensitivity analyses had broad similarity to the top module including *UBC*.

APP (chromosome 21: amyloid precursor protein) is another gene that appears to be important from these analyses. Its primary function is not known. It has been implicated as a regulator of synapse formation [29], neural plasticity [30], and iron export [31]. Although *APP* was the first gene identified for early onset familial AD, only recently has a relationship between *APP* and LOAD been reported [32,33]. These results are in contrast to other studies that have not found associations between LOAD and common [34] or rare [35] *APP* SNPs. One rare variant in *APP* was found to be protective for LOAD in Iceland [36];

this variant is rare in North Americans [37] and does not explain our findings, which are derived from variants with minor allele frequency >1%. Our finding that *APP* has central importance in the DMS-based top module reinforces the relevance of Aβ biology in LOAD pathogenesis.

EGR1 (early growth response protein 1; chromosome 5) has a distinct pattern of expression in brain, and its induction is associated with neuronal activity. *EGR1* regulates phosphorylation of microtubule-associated protein tau in mammalian brain [38], and *EGR1*-controlled regulatory networks are associated with neurodegeneration [39].

ALB (albumin: chromosome 4) is a soluble, monomeric protein which comprises about half of the blood serum protein. Albutein, a therapeutic albumin, was found to inhibit Aβ self-association by selectively binding Aβ aggregates and by preventing further growth of Aβ assemblies [40]. The Alzheimer's Disease Management by Albumin Replacement project found that therapeutic albumin was associated with mobilization of Aβ and cognitive improvement in treated patients [41].

ATP5H (chromosome 17) encodes subunit d of the enzyme mitochondrial ATP synthase [42,43]. In a recent article [44], a variant in the *ATP5H-KCTD2* locus was found to be associated with LOAD risk. Another mitochondrial gene, *NDUFS3*, which encodes complex I, mitochondrial respiratory chain, 30-kD subunit, also emerges from this systems-based approach. Knockdown of orthologs of both genes in *C. elegans* delayed paralysis because of Aβ$_{3-42}$ toxicity. Both genes had lower expression in the temporal cortex of people with LOAD compared with people with other neurodegenerative conditions and were significantly positively correlated with one another and with *UBC*.

For a complex disease such as LOAD, there may be a few rare variants with large effect size, and also multiple common variants, each with a more modest risk [45]. Our results suggest that genetic signals with modest association P values when considered independently (e.g., UBC, APP, EGR1, and ALB) could converge in interactome modules. These genes have weak independent association signals but are highlighted in the PPI analyses because of their extensive biological interactions with multiple additional genes that also had weak independent associations with LOAD risk when considered in isolation (one variant at a time). The PPI approach enables identification of an entire module of genes characterized by good evidence for relationships with each other and high representation of associations with the LOAD phenotype.

The gene-wide findings were mostly consistent with previous genetic analysis of LOAD with the same data [1,20]. The difference in our gene-wide results compared with the previously published results [20] can be attributed to the different mapping schemes used to link SNPs to genes. Using haplotype files for the 1000G reference build and a ±50 kb gene boundary, we were able to use 96% of the SNPs from the IGAP data, which resulted in more genes included in our analyses. Also, we used a different gene-wide analysis technique compared with that of Escott-Price et al. [20].

Our findings were different from those of network analyses using gene expression data, which identified an immune and microglia-specific module [3] and another study which also identified immune response, regulation of endocytosis, cholesterol transport, and protein ubiquitination pathways [4]. Those studies began with gene expression data, whereas our analyses began with SNP data. Those studies used curated pathways, whereas ours used a PPI approach. Those studies did not have functional validation, whereas we used a C. elegans Aβ proteotoxicity approach that richly confirmed the relevance of novel genes in our top module to Aβ biology in living animals.

Unlike some pathway approaches, the DMS approach does not require a priori curation of pathways. Instead, this method incorporates PPI data mined from publicly available databases to prioritize the genome-wide genetic data. Our results identified a top module that had biological plausibility, a high proportion of differentially expressed genes in the temporal cortex of people with LOAD compared with people with non-LOAD neurodegenerative conditions, and for four genes not previously associated with LOAD that had nematode orthologs, had functional outcomes in an aging-related nematode Aβ proteotoxicity models.

There are some limitations to our analyses that should be considered. As in any gene-wide analysis, although many SNPs are in genes or very close to genes and fall within our ±50 kb boundaries, some SNPs are outside those boundaries; some SNPs within those boundaries may be associated with the expression of a distal rather than the most proximal gene. PPIs may be tissue dependent. PPI databases document interactions between proteins that scientists have chosen to study and publish for more than the past 100 years. Almost certainly, additionally important interactions remain to be identified. Additional pharmacogenomics data may help disentangling the pathophysiologic implications of these genes.

Definitive nematode orthologs do not exist for most of the genes in the top module. We were thus unable to examine all interesting candidates in the nematode model. The C. elegans models of Aβ toxicity fail to recapitulate many of the important features of LOAD. Also, ubq-1 and ubq-2 shorten lifespan significantly, and there is a possibility that this knockdown creates a synthetic sick phenotype that accelerates paralysis in the context of $A\beta_{3-42}$ expression. Nevertheless, this model is useful for understanding genetic modifiers of cellular proteotoxic stress in a metazoan. Knockdown of ubq-1 and ubq-2 dramatically shortens lifespan, whereas knockdown of nuo-2, egrh-1, and atp-5 all extend lifespan. None of these effects are because of differences in development, as the data are shown as age in days of adulthood. It is not necessarily surprising that the loss of function in different network components could have opposing effects on protein homeostasis and Aβ toxicity in C. elegans. Functional validation in C. elegans is a way to demonstrate a role for these conserved factors in mediating Aβ toxicity, either through enhancing sensitivity or resistance. Positive C. elegans results (either reduced or enhanced resistance to Aβ toxicity) strongly suggest that the knocked out gene may be a conserved modifier of protein homeostasis. These findings should guide future mechanistic studies in C. elegans and mammalian systems.

It is of potential interest that knockdown of atp-5 and nuo-2 conferred protection against β-amyloid in C. elegans, whereas ATP5H and NDUFS3 expressions were reduced in the LOAD brains. One potential explanation for this could be that reduced expression of these, and perhaps other, electron transport chain components may be a protective response to accumulation of Aβ in the brain. Future studies in C. elegans may shed light on the mechanisms by which knockdown of these mitochondrial proteins enhances resistance to Aβ and whether a similar decrease in the expression of these genes is associated with transgenic expression of Aβ in C. elegans.

The brain expression data we analyzed here were derived from people with pathologically confirmed LOAD and people with other non-LOAD related neurodegenerative conditions. Identifying differences in gene expression between these groups can indicate genes that may play a role in LOAD pathogenesis. However, these data do not inform us as to whether there may be differences in the expression for these genes between individuals with LOAD and those without any neurodegenerative conditions [16]. Except for ALB expression, we did not find any differences between cerebellar expression levels of people with LOAD and people with other causes of neurodegeneration for the genes we identified in our top module. Because temporal cortex is a

region that is significantly affected with LOAD neuropathology, changes in gene expression detected within the temporal cortex and not in cerebellum may be a consequence rather than a cause of the pathology. We tried to control for cellular loss (i.e., neurons) or increase (i.e., glia) in LOAD temporal cortex by including cell-specific probes in our analyses of gene expression, but this approach may not be sufficient to account for all neuropathology-driven expression differences. Although it is not possible to discern whether gene expression differences are a cause or consequence of neuropathology using expression profiling, this approach can nevertheless identify genes that are key in disease pathophysiology (e.g., *APP*) and also provide important, additional evidence for genes implicated in a disease by other approaches (i.e., gene-based association, interactome, and so forth). We note that our microarray-based measurements of expression levels from tissue cannot discern expression changes obtained at the single cell or cell-component level. Additional studies, such as [46], are needed to determine, for example, whether these expression differences are driven by specific cell types within the same tissue, or within neurons whether these differences are driven by cell body versus dendritic versus synaptic domains, which have been shown to exist for some of the proteins reported here.

In summary, we used a DMS approach to identify modules of genes associated with LOAD. We confirmed some prior findings that used complementary analytic strategies. We also identified some loci not previously associated with LOAD. We used *C. elegans* models to confirm Aβ-related proteotoxicity associated with four of these novel loci and found enrichment for differentially expressed genes in the temporal cortex from people with LOAD compared with people with non-LOAD neurodegeneration. Subsequent analyses may identify therapeutic targets associated with these loci.

Acknowledgments

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jalz.2017.01.016.

RESEARCH IN CONTEXT

1. Systematic review: We searched for "protein-protein interaction [1]" (PPI), "SNP" and "Alzheimer's disease" (AD) in PubMED and identified an article on July 28, 2016 where the authors used a gene-gene core-regulation network based on *cis*-expression quantitative trait loci (eQTL) SNPs and a single PPI database. This study represents the most comprehensive PPI-based network analyses for AD integrating all SNPs from the biggest genome-wide association study of AD [2] and PPI mined from 11 databases.

2. Interpretation: The study demonstrates use of a novel approach to prioritize genetic association results by integrating prior biological knowledge. RNA interference knockdowns of the *Caenorhabditis elegans* orthologs of *UBC, NDUFS3, EGR1*, and *ATP5H* were significantly associated with Aβ toxicity, and *NDUFS3, SLC25A11, ATP5H*, and *APP* were differentially expressed in the temporal cortex.

3. Future directions: Genes that may not be identified in standard genome-wide association study analyses may play an important role in the pathophysiology of late-onset AD.

References

[1] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013; 45:1452–8.

[2] Peng G, Luo L, Siu H, Zhu Y, Hu P, Hong S, et al. Gene and pathway-based second-wave analysis of genome-wide association studies. Eur J Hum Genet 2010;18:111–7.

[3] Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 2013;153:707–20.

[4] International Genomics of Alzheimer's Disease Consortium (IGAP). Convergent genetic and expression data implicate immunity in Alzheimer's disease. Alzheimers Dement 2015;11:658–71.

[5] Liang H, Li WH. MicroRNA regulation of human protein protein interaction network. RNA 2007;13:1402–8.

[6] Akula N, Baranova A, Seto D, Solka J, Nalls MA, Singleton A, et al. A network-based approach to prioritize results from genome-wide association studies. PLoS One 2011;6:e24220.

[7] Liu JZ, Mcrae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet 2010;87:139–45.

[8] Chanda P, Huang H, Arking DE, Bader JS. Fast association tests for genes with FAST. PLoS One 2013;8:e68585.

[9] Mora A, Donaldson IM. iRefR: An R package to manipulate the iRefIndex consolidated protein interaction database. BMC Bioinformatics 2011;12:455.

[10] Jia PL, Zheng SY, Long JR, Zheng W, Zhao ZM. dmGWAS: Dense module searching for genome-wide association studies in protein-protein interaction networks. Bioinformatics 2011;27:95–102.

[11] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.

[12] Huang H, Chanda P, Alonso A, Bader JS, Arking DE. Gene-based tests of association. PLoS Genet 2011;7:e1002177.

[13] Link CD. Expression of human beta-amyloid peptide in transgenic *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 1995; 92:9368–72.

[14] McColl G, Roberts BR, Pukala TL, Kenche VB, Roberts CM, Link CD, et al. Utility of an improved model of amyloid-beta (Abeta(1)(-)(4)(2)) toxicity in *Caenorhabditis elegans* for drug screening for Alzheimer's disease. Mol Neurodegener 2012;7:57.

[15] Steinkraus KA, Smith ED, Davis C, Carr D, Pendergrass WR, Sutphin GL, et al. Dietary restriction suppresses proteotoxicity and enhances longevity by an hsf-1-dependent mechanism in *Caenorhabditis elegans*. Aging Cell 2008;7:394–404.

[16] Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS Genet 2012;8:e1002707.

[17] Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. Nat Genet 2009;41:192–8.

[18] Mukherjee S, Kim S, Gibbons L, Nho K, Risacher S, Glymour M, et al. Genetic architecture of resilience of executive functioning. Brain Imaging Behav 2012;6:621–33.

[19] Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003;100:9440–5.

[20] Escott-Price V, Bellenguez C, Wang LS, Choi SH, Harold D, Jones L, et al. Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. PLoS One 2014;9:e94661.

[21] Jin K, Mao XO, Sun Y, Xie L, Jin L, Nishi E, et al. Heparin-binding epidermal growth factor-like growth factor: hypoxia-inducible expression in vitro and stimulation of neurogenesis in vitro and in vivo. J Neurosci 2002;22:5365–73.

[22] Lipovich L, Lynch ED, Lee MK, King MC. A novel sodium bicarbonate cotransporter-like gene in an ancient duplicated region: SLC4A9 at 5q31. Genome Biol 2001;2. RESEARCH0011.

[23] International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin UM, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet 2011;377:641–9.

[24] Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet 2010; 42:781–5.

[25] International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, et al. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med 2007;357:851–62.

[26] Mero IL, Gustavsen MW, Saether HS, Flam ST, Berg-Hansen P, Sondergaard HB, et al. Oligoclonal band status in Scandinavian multiple sclerosis patients is associated with specific genetic risk alleles. PLoS One 2013;8:e58352.

[27] Ferrari R, Hernandez DG, Nalls MA, Rohrer JD, Ramasamy A, Kwok JB, et al. Frontotemporal dementia and its subtypes: A genome-wide association study. Lancet Neurol 2014;13:686–99.

[28] Atkin G, Paulson H. Ubiquitin pathways in neurodegenerative disease. Front Mol Neurosci 2014;7:63.

[29] Priller C, Bauer T, Mitteregger G, Krebs B, Kretzschmar HA, Herms J. Synapse formation and function is modulated by the amyloid precursor protein. J Neurosci 2006;26:7212–21.

[30] Turner PR, O'Connor K, Tate WP, Abraham WC. Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory. Prog Neurobiol 2003;70:1–32.

[31] Duce JA, Tsatsanis A, Cater MA, James SA, Robb E, Wikhe K, et al. Iron-export ferroxidase activity of beta-amyloid precursor protein is inhibited by zinc in Alzheimer's disease. Cell 2010;142:857–67.

[32] Rosenthal SL, Kamboh MI. Late-onset Alzheimer's disease genes and the potentially implicated pathways. Curr Genet Med Rep 2014; 2:85–101.

[33] Kero M, Paetau A, Polvikoski T, Tanskanen M, Sulkava R, Jansson L, et al. Amyloid precursor protein (APP) A673T mutation in the elderly Finnish population. Neurobiol Aging 2013; 34:1518.e1–1518.e3.

[34] Nowotny P, Simcock X, Bertelsen S, Hinrichs AL, Kauwe JS, Mayo K, et al. Association studies testing for risk for late-onset Alzheimer's disease with common variants in the beta-amyloid precursor protein (APP). Am J Med Genet B Neuropsychiatr Genet 2007;144B:469–74.

[35] Sassi C, Guerreiro R, Gibbs R, Ding J, Lupton MK, Troakes C, et al. Investigating the role of rare coding variability in Mendelian dementia genes (APP, PSEN1, PSEN2, GRN, MAPT, and PRNP) in late-onset Alzheimer's disease. Neurobiol Aging 2014;35:2881.e1–2881.e6.

[36] Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 2012;488:96–9.

[37] Bamne MN, Demirci FY, Berman S, Snitz BE, Rosenthal SL, Wang X, et al. Investigation of an amyloid precursor protein protective mutation (A673T) in a North American case-control sample of late-onset Alzheimer's disease. Neurobiol Aging 2014;35:1779.e15–1779.e16.

[38] Lu Y, Li T, Qureshi HY, Han D, Paudel HK. Early growth response 1 (Egr-1) regulates phosphorylation of microtubule-associated protein tau in mammalian brain. J Biol Chem 2011;286:20569–81.

[39] Koldamova R, Schug J, Lefterova M, Cronican AA, Fitz NF, Davenport FA, et al. Genome-wide approaches reveal EGR1-controlled regulatory networks associated with neurodegeneration. Neurobiol Dis 2014;63:107–14.

[40] Milojevic J, Costa M, Ortiz AM, Jorquera JI, Melacini G. In vitro amyloid-beta binding and inhibition of amyloid-beta self-association by therapeutic albumin. J Alzheimers Dis 2014;38:753–65.

[41] Boada M, Ramos-Fernandez E, Guivernau B, Munoz FJ, Costa M, Ortiz AM, et al. Treatment of Alzheimer disease using combination therapy with plasma exchange and haemapheresis with albumin and intravenous immunoglobulin: Rationale and treatment approach of the AMBAR (Alzheimer Management By Albumin Replacement) study. Neurologia 2016;31:473–81.

[42] Kinosita K Jr, Yasuda R, Noji H. F1-ATPase: A highly efficient rotary ATP machine. Essays Biochem 2000;35:3–18.

[43] Oster G, Wang H. Rotary protein motors. Trends Cell Biol 2003; 13:114–21.

[44] Boada M, Antunez C, Ramirez-Lorca R, DeStefano AL, Gonzalez-Perez A, Gayan J, et al. ATP5H/KCTD2 locus is associated with Alzheimer's disease risk. Mol Psychiatry 2014; 19:682–7.

[45] Ridge PG, Mukherjee S, Crane PK, Kauwe JS. Alzheimer's Disease Genetics Consortium. Alzheimer's disease: analyzing the missing heritability. PLoS One 2013;8:e79771.

[46] Valla J, Berndt JD, Gonzalez-Lima F. Energy hypometabolism in posterior cingulate cortex of Alzheimer's patients: superficial laminar cytochrome oxidase associated with disease duration. J Neurosci 2001;21:4923–30.