Integrated Whole Transcriptome and DNA Methylation Analysis Identifies Gene Networks Specific to Late-Onset Alzheimer's Disease

Crystal E. Humphries[a], Martin A. Kohli[b], Lubov Nathanson[c,e], Patrice Whitehead[b], Gary Beecham[a,b], Eden Martin[a,b], Deborah C. Mash[d], Margaret A. Pericak-Vance[a,b] and John Gilbert[a,b,*]

[a] University of Miami, Miller School of Medicine, Dr. John T. Macdonald Foundation Department of Human Genetics, Miami, FL, USA
[b] University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics, Miami, FL, USA
[c] University of Miami, Miller School of Medicine, Department of Medicine, Miami, FL, USA
[d] University of Miami, Miller School of Medicine, Miami Brain Bank Endowment, FL, USA
[e] Nova Southeastern University, Ft. Lauderdale, FL, USA

Abstract. Previous transcriptome studies observed disrupted cellular processes in late-onset Alzheimer's disease (LOAD), yet it is unclear whether these changes are specific to LOAD, or are common to general neurodegeneration. In this study, we address this question by examining transcription in LOAD and comparing it to cognitively normal controls and a cohort of "disease controls." Differential transcription was examined using RNA-seq, which allows for the examination of protein coding genes, non-coding RNAs, and splicing. Significant transcription differences specific to LOAD were observed in five genes: *C10orf105*, *DIO2*, a lincRNA, *RARRES3*, and *WIF1*. These findings were replicated in two independent publicly available microarray data sets. Network analyses, performed on 2,504 genes with moderate transcription differences in LOAD, reveal that these genes aggregate into seven networks. Two networks involved in myelination and innate immune response specifically correlated to LOAD. *FRMD4B* and *ST18*, hub genes within the myelination network, were previously implicated in LOAD. Of the five significant genes, *WIF1* and *RARRES3* are directly implicated in the myelination process; the other three genes are located within the network. LOAD specific changes in DNA methylation were located throughout the genome and substantial changes in methylation were identified within the myelination network. Splicing differences specific to LOAD were observed across the genome and were decreased in all seven networks. DNA methylation had reduced influence on transcription within LOAD in the myelination network when compared to both controls. These results hint at the molecular underpinnings of LOAD and indicate several key processes, genes, and networks specific to the disease.

Keywords: Alzheimer's disease, dementia with Lewy bodies, DNA methylation, expression, human brain, Illumina Human Methylation 450K, late-onset transcriptome, myelination, networks, RNA-seq, splicing

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia affecting approximately 13% of individuals 65 and older [1]. Rare variants in the amyloid precursor protein, presenilin 1, or presenilin 2 genes

*Correspondence to: John Gilbert, PhD, University of Miami Miller School of Medicine, 1501 NW 10th Avenue, Biomedical Research Building, Suite 616, Miami, FL 33136, USA. Tel.: +1 305 243 2282; Fax: +1 305 243 2703; E-mail: jgilbert@med.miami.edu.

primarily cause early-onset AD and comprise 2% of AD cases [2–5]. Conversely, late-onset AD (LOAD) is much more complex and accounts for greater than 90% of AD cases [6].

Recent transcriptome studies have shed light on molecular functions altered in LOAD. Many of these studies have consistently observed alterations within several pathways: synaptic transmission [7, 8], inflammation [9], energy metabolism [9–11], myelination [11], and signal transduction [8, 10, 12]. While amyloid-β (Aβ) plaques and neurofibrillary tangles (NFTs) are pathological hallmarks of AD, many clinical LOAD phenotypes are shared across other dementias, including cognitive decline and memory loss [13–15]. Transcriptome studies have predominately compared AD samples to normal controls [16]. Using this approach, it is often unclear if the observed disrupted processes in LOAD are specific to LOAD or are the result of non-specific general neurodegenerative effects.

This study approaches the problem of distinguishing the processes altered specifically in LOAD from those generally involved in neurodegeneration. To do this, we performed whole transcriptome analysis on temporal pole brain tissue (Brodmann's Area 38) from three cohorts: patients with LOAD, cognitively normal controls, and patients with dementia with Lewy bodies (DLB) which we term "disease controls." While no disease control is perfect, DLB samples are excellent disease controls due to both diseases sharing clinical characteristics, overlapping affected brain regions, and "extracellular body" pathologies. That is, LOAD has Aβ plaques and DLB has Lewy bodies. Similarly, some of the affected regions, such as the temporal pole region examined in this study, overlap extensively. The disease controls allow us to attempt to decipher processes and genes altered as a result of general dementia from those specific to the LOAD process.

In this study, we identified LOAD specific transcriptional changes of protein coding and non-coding RNAs by comparing transcription in LOAD to cognitively normal controls and disease controls. To identify transcriptional differences, we utilized RNA-seq technology. RNA-seq allows for the detection of known and novel isoforms of well-documented genes as well as non-coding RNAs. Examination of individual genes and network analysis demonstrates that LOAD specific transcriptional differences converge upon several processes including innate immune response and myelination. To determine if this transcriptional process was disrupted due to altered epigenetic regulation, we examined DNA methylation and splicing. In this study, we identify differentially transcribed and spliced genes within networks and DNA methylation changes that correlate specifically to LOAD.

MATERIAL AND METHODS

Samples

RNA transcription was investigated using tissue samples isolated from the temporal pole of a total of thirty brain samples. Ten samples were collected from each of the following three cohorts: subjects with LOAD, cognitively normal controls, and disease controls, subjects with DLB. Samples were extracted from the temporal pole (Brodmann area 38) of age-matched Caucasian males (Supplementary Table 1). The mean (SD) ages were LOAD: 77.4 ($\pm$5.7) years; DLB: 79.1 ($\pm$5.6) years; cognitively normal controls: 74.6 ($\pm$7.8) years. Samples were frozen and stored at -80°C.

Tissue

All cases underwent a standardized neuropathological assessment with evaluation of gross and microscopic findings and quantitative analysis of Alzheimer-type pathology. LOAD cases were selected according to dementia status, staged for LOAD pathology according to Braak (III, IV), and were positive for Aβ and PHF-tau in two brain areas (Brodmann areas 9 and 39) [16].

Semi-quantitative grading of Lewy body pathology and assignment of Lewy body type were determined according to the Third CDLB recommendations [17, 18]. Cases of DLB were selected based on the distribution of Lewy bodies and the severity of Alzheimer-type pathology. The DLB cases included in the study had a more extensive neocortical type of distribution that included the temporal lobe. All of the transitional and diffuse neocortical cases were from demented subjects. Based on the likelihood categories [17, 18], the extensive alpha synucleinopathy corresponds to an intermediate or high likelihood of DLB in this cohort.

Normal control samples were confirmed to be cognitively normal and died from causes unrelated to neurological disease. Autopsies and neuropathologic diagnoses were performed in accordance with published guidelines by a consultant neuropathologist. Extracted tissue included both grey and white matter.

RNA and DNA isolation

RNA was isolated from frozen tissue samples using the QIAGEN Qiashredder and TRIzol reagents

(Invitrogen). RNA was extracted using the miRNeasy mini kit (Invitrogen) and each sample was treated with an on-column DNAse treatment (Invitrogen). RNA was dissolved in DNase/RNase-free water (Invitrogen) and concentration was determined using the Qubit Fluorometer with the Qubit™ RNA kit. The quality of the RNA was determined using the Bioanalyzer 2100 (Agilent).

For methylation analysis, genomic DNA was isolated from 24 of the 30 frozen tissue samples used for transcription studies: eight LOAD patients, eight cognitively normal controls, and eight DLB patients. Isolation was carried out using the QIAamp® DNA Mini and Blood Mini Handbook (Qiagen) in accordance with the manufacturer's specifications. DNA concentration was determined using the Qubit Fluorometer with the Qubit™ DNA Broad Range kit. DNA integrity was assessed using gel electrophoresis. All RNA and DNA samples were stored at –80°C.

RNA-seq library preparation

RNA-seq libraries were prepared from 10 μg of total RNA isolated from each sample. The RNA integrity number (RIN) was determined using the Bioanalyzer 2100, and all samples had a RIN number ≥6 (Supplementary Table 1). Ribosomal RNA was depleted using the RiboMinus™ Eukaryote Kit for RNA-seq (Invitrogen). Depletion was confirmed using the Agilent RNA Nano chip and Bioanalyzer 2100. RNA concentrations were determined using the Qubit™ RNA kit. Approximately 600 ng of ribosomal depleted RNA was utilized for library preparation using Script-Seq (Epicentre®) along with the Phusion® Polymerase enzyme (Kappa Biosystems). Library completion was confirmed using the DNA High Sensitivity Kit on the Bioanalzyer 2100 and concentration was determined using the Library Quantification Kit-Illumina (Kappa Biosystems).

DNA methylation library preparation

Bisulfite conversion of 500 ng of genomic DNA was achieved with the EZ DNA Methylation Gold kit (Zymo Research). DNA samples were prepared according to the Illumina©Infinium protocol and run on the Illumina©Infinium HumanMethylation450 bead chip.

RNA-seq analysis

Libraries were sequenced on the Illumina HiSeq2000 with an average of 50 million 100 bp paired-end reads being sequenced per library. Reads were aligned using GSNAP software and only unique reads were used in analyses [19]. To ensure the greatest percentage of aligned reads, bar codes were trimmed off the sequencing read prior to alignment. Strandedness was assigned based on the Script-seq library protocol.

To examine transcription and splicing differences, reads were assembled to transcripts from the Gencode v15 database to generate count data for each transcript using SAMtools [20]. To reduce noise and increase reliability, only transcripts with a count above five were used in subsequent analysis. Transcripts were normalized and differences were determined using the DESeq v1.12.0 software [21]. Splicing differences were resolved using the DEXSeq v1.12.0 software [22]. To accurately assess splicing, only genes four or more exons were examined, which resulted in a total of 17,076 genes evaluated for splicing differences. Both of these analyses, as well as downstream analyses, were carried out using R software, version 3.0.1 (http://www.r-project.org). This data will be publicly available in the Sequence Read Archive (SRA) database.

DNA methylation analysis

DNA methylation was performed using the ChAMP software available from bioconductor [23]. All default parameters were used except that the β-value, which is the ratio of methylation probes to both un-methylated and methylated probes, was beta-quantile normalized. Using the package Limma, a generalized linear model with disease controls as a covariate, CpG sites were assessed for differential methylation [24]. This data will be publicly available in the Gene Expression Omnibus (GEO) database.

Network analysis

Using Weighted Gene Correlation Network Analysis (WGCNA), we examined potential networks, as defined by genes' co-transcription relationships to each other [25, 26]. Co-expressed genes can denote potential functional relationships, and are used to identify hub gene(s) that are central to a network's function [26]. For analysis, we used genes with nominally significant ($p < 0.05$) differences in expression between LOAD cases and cognitively normal controls (total genes: 2,504). Network visualization was performed using VisAnt [27]. Pathway analysis was conducted using the Database for Annotation, Visualization and

Integrated Discovery (DAVID) [28]. Correlations of networks between LOAD and DLB were analyzed using Dunn and Clark on the Cocor R-package [29, 30].

Validation by independent data set replication

To replicate transcription results we used two separate publically available data sets. In the first data set, cortex samples from 363 individuals (GSE155222) were assessed for differential transcription [31]. This sample consisted of 363 individuals: 176 normal controls and 183 LOAD samples (Supplementary Table 1). Initially, we adjusted for confounders to account for age, gender, and gender variables on transcription. To do this, we used a robust linear regression model for covariate corrections as equation (2). The residuals were used for further differential transcription. This analysis was performed using the MASS packages [32]

$$(2) \quad transcription = \beta_1 \, PMI + \beta_2 \, age + \beta_3 \, gender$$

In the second data set, which is completely independent from the first data set, prefrontal cortex samples from 227 individuals (GSE44770) were assessed for differential transcription [33]. This sample consisted of 127 normal controls and 100 AD samples. Transcription values given were already normalized for age, gender, batch, and RIN. For both data sets, each gene was assessed for differential transcription between two conditions using a two-tailed *t*-test with a *p*-value threshold of 0.05. All *p*-values were corrected using Benjamini–Hochberg False Discovery Rate (FDR) and this analysis was performed using the multtest R-package [34].

RESULTS

To examine transcriptional differences in LOAD, we identified a total of 53,245 genes (19,207 protein-coding and 34,038 ncRNAs, based on ENCODEv15) having detectable levels of transcription [35]. Transcription of protein coding genes and each type of ncRNA [36] was similar between cohorts (Supplementary Fig. 1). While more ncRNA transcripts were observed than protein-coding transcripts, the average read depths of ncRNAs were less than protein-coding genes; a result observed in previously published whole-transcriptome studies [37].

Five genes with altered transcription specific to LOAD

Using DESeq [21], we examined individual genes for differences in transcription. A total of 2,504 out of 53,245 genes had nominal differences in transcription ($p < 0.05$) between the ten LOAD samples and ten normal controls (Supplementary Fig. 2). After correcting for multiple testing by False Discovery Rate (FDR <0.05), 16 of the 2,504 genes differed significantly between LOAD and normal controls. These genes consisted of eleven protein-coding genes, four long intergenic ncRNAs (lincRNAs), and a pseudogene (Table 1a).

To test whether differential transcription of these sixteen genes is LOAD specific relative to our disease control, we compared transcription of the genes to the ten disease controls. Five of the sixteen (four protein coding and one ncRNA) were still significant after FDR correction (Table 1b). Two genes were increased in LOAD (*C10orf105* and *RARRES3*), while three genes were decreased in LOAD (*DIO2, ENSG00000249343.1,* and *WIF1*).

These transcriptional findings were corroborated by two independent, publicly available, microarray datasets. The first dataset (GSE15522) examined cortex transcription of 363 individuals and the second dataset (GSE44770) examined transcription in the prefrontal cortex of 227 individuals [31, 33]. Independent analysis of each dataset yielded similar findings with both datasets having a significant decrease of both *WIF1* (Wnt-Inhibitory Factor 1) and *DIO2* (Type II Iodothyronine Deiodinase) in LOAD, whereas *RARRES3* (Retinoic Acid Receptor Responder 3) was significantly increased in LOAD (Fig. 1). *C10orf105* was only detectable in the second dataset and its transcription was similar to our findings in that transcription was significantly increased in LOAD when compared to controls. *ENSG00000249343.1* could not be verified due to the lack of probes for this gene in both microarray datasets.

Network analysis reveals myelination specific to LOAD

To gain insight into the processes disrupted in LOAD, we performed network analysis on the 2,504 genes that had nominal significance between LOAD and normal controls. Using Weighted Gene Co-transcription Network Analysis (*WGCNA*) [25] on transcription values of the 2,504 genes revealed that these genes aggregated into seven networks (Fig. 2). Each network is referred to by number because not all predicted networks have a unifying function [26]. We examined networks altered specifically in AD. We also investigated the networks altered as a result of the general neurodegenerative processes by

Table 1
Transcriptional differences specific to LOAD

A

Gene Symbol	Gene Name	Log$_2$FC	FDR	Biotype
ADAMTS2	ADAM Metallopeptidase with Thrombospondin Type 1 Motif, 2	1.93	0.009	Protein Coding
NRN1	Neuritin 1	−1.285	0.009	Protein Coding
WIF1	Wnt Inhibitory Factor 1	−2.015	0.009	Protein Coding
C10orf105	Chromosome 10 Open Reading Frame 105	2.105	0.009	Protein Coding
CDR1	Cerebellar Degeneration-Related Protein 1	1.933	0.013	Protein Coding
KCNV1	Potassium Channel, Subfamily V, Member 1	−1.293	0.014	Protein Coding
DIO2	Deiodinase, Iodothyronine, Type II	−1.180	0.014	Protein Coding
RP5-965F6.2	RP5-965F6.2	1.185	0.014	LincRNA
RP11-511P7.2	RP11-511P7.2	−2.987	0.014	LincRNA
IFLTD1	Intermediate Filament Tail Domain Containing 1	2.158	0.025	Protein Coding
RP11-318G8.1	RP11-318G8.1	2.158	0.025	Pseudogene
SCG2	Secretogranin II	−1.333	0.032	Protein Coding
CTD-2275D24.2	CTD-2275D24.2	−1.510	0.032	LincRNA
TXNIP	Thioredoxin Interacting Protein	1.845	0.033	Protein Coding
RARRES3	Retinoic Acid Receptor Responder (Tazarotene Induced) 3	1.601	0.033	Protein Coding
CTD-2378E21.1	CTD-2378E21.1	−2.275	0.037	LincRNA

B

CTD-2275D24.2	CTD-2275D24.2	−1.419	3.260 E-05	LincRNA
WIF1	Wnt Inhibitory Factor 1	−1.165	0.012	Protein Coding
C10orf105	Chromosome 10 Open Reading Frame 105	1.319	0.015	Protein Coding
RARRES3	Retinoic Acid Receptor Responder (Tazarotene Induced) 3	0.742	0.066	Protein Coding
DIO2	Deiodinase, Iodothyronine, Type II	−0.694	0.081	Protein Coding

Transcription was compared between LOAD and cognitively normal controls that showed sixteen transcripts to be significant with a false discovery rate (FDR) below 0.05 (A). The sixteen genes were then compared between LOAD and disease controls. Five genes were significant with an FDR below 0.1 (B).

examining networks that were correlated to both LOAD and the disease controls.

Five of the seven networks are correlated to both AD and disease controls relative to normal controls (Fig. 2a). Of these five that correlate to both LOAD and DLB, network 3 had the strongest association with LOAD and DLB ($r = 0.85$) followed by networks 4, 5, 6, and 7 (Fig. 2a). Network 3 was enriched for
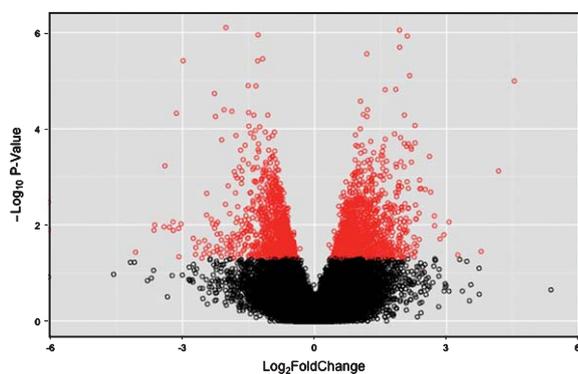

Fig. 1. Differential transcription in LOAD versus normal controls. Volcano plot for differential gene transcription between 10 cognitively normal controls and 10 LOAD samples. The most significant genes with differential transcription are shown in red. There are slightly more genes that are increased in expression in LOAD.

genes involved in behavior networks while networks 4, 5, 6, and 7 were enriched for generation of precursor metabolites, synaptic transmission, calcium ion binding, and immune responses, respectively (Supplementary Table 2).

Two of the seven networks were solely associated with LOAD, suggesting these networks are specific to AD (Supplementary Fig. 3). One of the LOAD specific networks, network 2, contained 42 genes and was enriched for GO pathways involved in innate immune responses (Benjamini–Hochberg (BH) p-value = 9.5e-6, see Supplementary Table 2). Network 1 contained 465 genes and had the highest correlation with LOAD ($r = 0.62$). It was primarily enriched in genes involved in myelination (BH p-value = 5.3e-3, see Supplementary Table 2). Hub genes are highly connected and are likely to be instrumental in the organization of a network [38]. Our findings showed ST18 and *FRMD4B* to be two prominent hub genes within the myelination network (Fig. 2b). These two genes have also been observed in prior studies to be hub genes of myelination networks disrupted in LOAD [33, 39]. Of the five genes with altered transcription in LOAD, the *C10orf105, RARRES3,* and *ENSG0000249343.1* genes are all located in the myelination network detected above and *WIF1* and *DIO2* are functionally related to the myelination process (see Discussion).
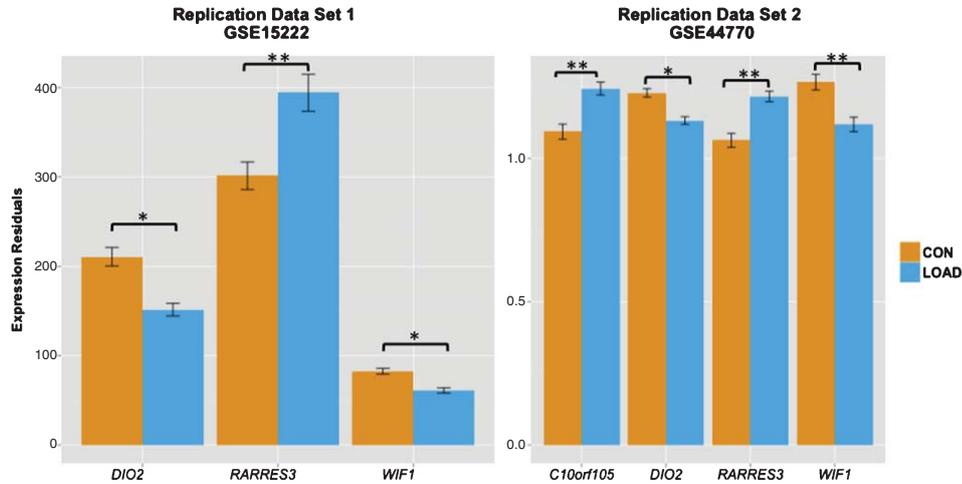
Fig. 2. Transcriptional differences in LOAD are replicated in two independent data sets. *DIO2, RARRES3,* and *WIF1* findings were replicated in the *GSE15222* data set which consists of 363 cortex samples. *C10orf105, DIO2, RARRES3,* and *WIF1* were similarly replicated in the GSE44770 data set, which consisted of 227 pre-frontal cortex samples. Each gene was assessed for differential transcription between LOAD and CON using a two-tailed *t*-test with a *p*-value threshold of 0.05. *p*-values for all genes were below 0.01. Both independent data sets consisted of cohorts of Caucasian descent.

Further examination of transcription in the myelination network revealed that 428 of the 465 genes have increased transcription in LOAD (Supplementary Table 3). We next examined methylation to determine if the transcriptional changes in these network genes correlated to changes in DNA methylation. Using the Infinium HumanMethylation450 bead array, we identified 5,147 CpG sites that overlapped the 465 genes in the myelination network. A total of 1,106 of the 5,147 CpG sites differed between LOAD and both controls ($p \leq 0.05$) (Table 2). The amount of altered methylation within the myelination network was significantly enriched when compared to the other networks (χ^2; $p < 1.0\text{e-}8$). Hypomethylation in LOAD was observed in 87.3% of the 1,106 CpG sites within the myelination network (Fig. 3). Similar changes were not observed in the innate immune network.

DNA methylation located within the gene body is also known to have a positive correlation with splicing [40]. To determine if there are splicing changes in the genes within the myelination network, we examined differential exon usage (DEU) between LOAD and both controls [22]. DEU in LOAD was observed in over 25% of genes when compared to both controls. This finding confirms an earlier study that also examined LOAD using RNA-seq [8]. Splicing within the myelination pathway was lower than expected. Our findings show that only 32% of the genes in the myelination network have DEU, whereas 37% of genes within all network genes have DEU (Supplementary Table 5). These results suggest that hypomethylation within the gene body influences the decreased splicing observed in the myelination network.

Table 2
Hypo/Hyper-methylated CpG sites overlapping genes within the networks

Network	Differentially methylated CpG* Total (%)	Hypermethylated CpGs + (Fold-Change)	Hypomethylated CpGs − (Fold-Change)
Network 1	1106/5147 (22%)	141	965
Network 2	81/431 (18%)	19	62
Network 3	414/2169 (19%)	89	325
Network 4	1236/7392 (17%)	601	635
Network 5	559/3765 (14%)	280	279
Network 6	704/3843 (18%)	134	570
Network 7	666/3540 (18%)	99	567

Genes with differential methylated CpG sites within genes in the networks are listed in the "Differentially methylated CpG" column along with the total number of CpG sites detected within the network. The number of differential hypomethylated and hypermethylated sites is listed for each network.
*Differentially methylated CpG sites had a value of $p < 0.05$.

A

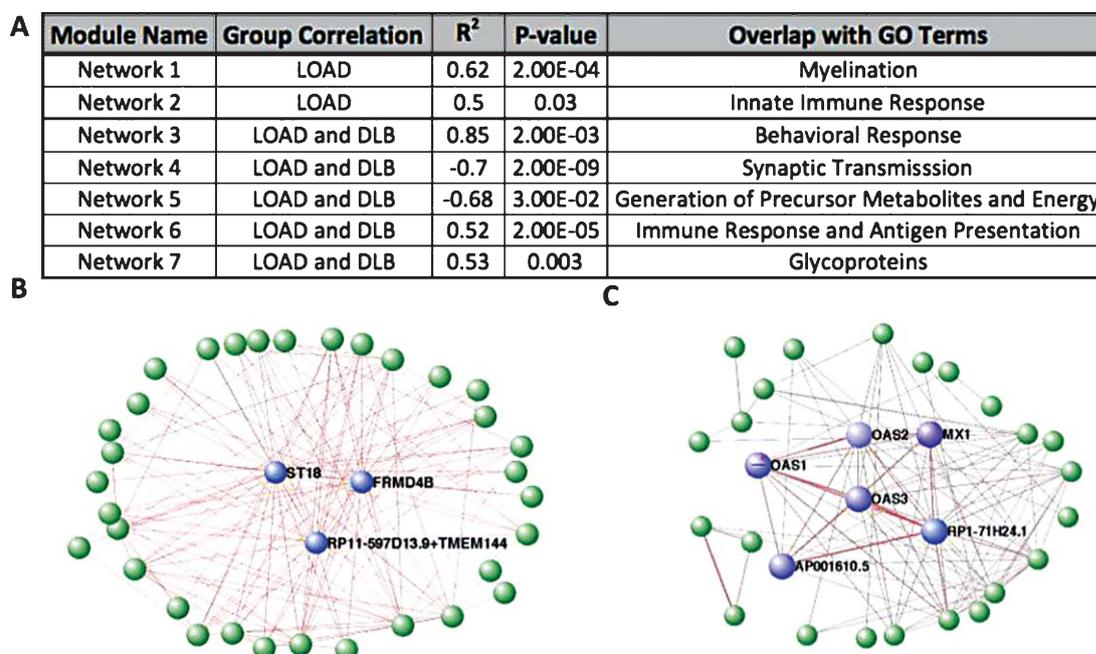| Module Name	Group Correlation	R^2	P-value	Overlap with GO Terms
Network 1	LOAD	0.62	2.00E-04	Myelination
Network 2	LOAD	0.5	0.03	Innate Immune Response
Network 3	LOAD and DLB	0.85	2.00E-03	Behavioral Response
Network 4	LOAD and DLB	-0.7	2.00E-09	Synaptic Transmisssion
Network 5	LOAD and DLB	-0.68	3.00E-02	Generation of Precursor Metabolites and Energy
Network 6	LOAD and DLB	0.52	2.00E-05	Immune Response and Antigen Presentation
Network 7	LOAD and DLB	0.53	0.003	Glycoproteins

B   **C**

Fig. 3. Networks specific to LOAD and general neurodegeneration. The 2,504 genes with differential transcription between LOAD and normal controls aggregated into seven networks (A). Two of these networks were specific to LOAD: (B) Network 1 and (C) Network 2. The hub genes of each network are labeled and colored blue. Each hub is connected to >20 genes with an $r \geq 0.35$. The hub genes in network 2 have at least 10 genes connected to them with an $r \geq 0.15$.
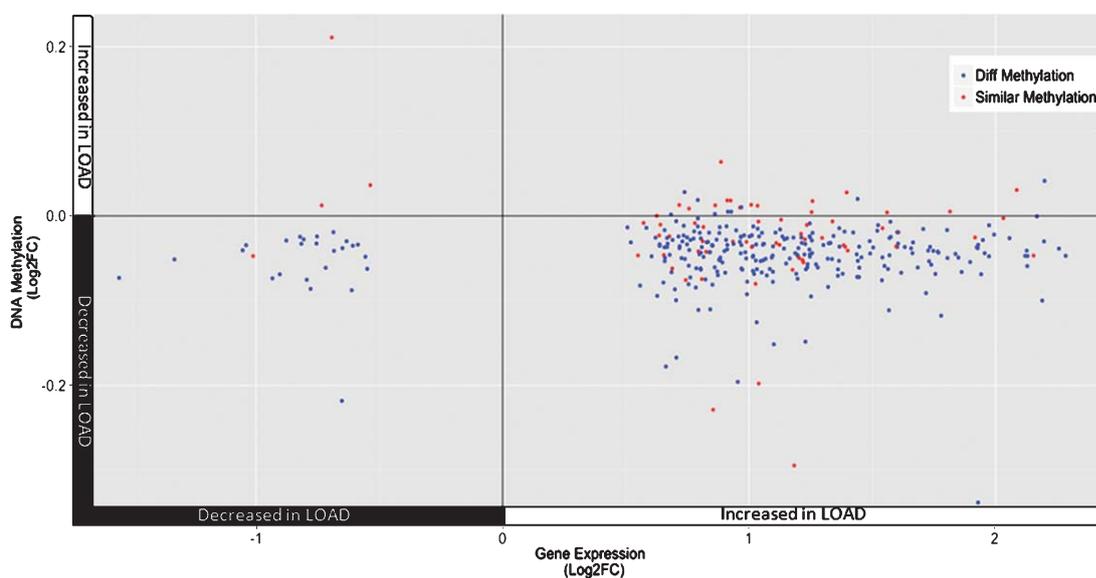

Fig. 4. Hypomethylation in the myelination network. The log 2 fold-change of gene transcription, when comparing LOAD to CON, is plotted against the log 2 fold-change of DNA methylation averaged across the gene (LOAD versus CON+D.CON). Most of the genes in the network have decreased methylation and increased transcription (located in the lower right quadrant). Genes with a CpG site that is differentially methylated are colored blue.

DISCUSSION

Previous transcriptome studies identified pathways altered in LOAD either as a specific consequence of disease or as a result of the effects of general neurodegeneration. While our initial findings included finding genes that were previously found to be altered in LOAD (Supplementary Tables 5 and 6), including disease controls allowed us to separate processes disrupted during the neurodegenerative process from those disrupted as a direct consequence of LOAD [41].

There are similarities in the presentations of AD and DLB. AD and DLB are diseases that affect memory and cause changes in cognition and reasoning [42]. Regions, such as the temporal lobe, which have Aβ plaques and NFTs in AD overlap with some of the regions affected in DLB [43–46]. The onset of DLB is typically 60 to 85 years of age, similar to LOAD [47]. Clinically, the two diseases are often indistinguishable; however, their differences are apparent during postmortem neuropathological examination. Pathologically, DLB can be identified by the accumulation of foreign bodies, such as α-synuclein aggregates and NFTs, within the brain [48]. DLB has a more rapid progression and causes a decrease in motor function [42]. While no control is perfect, DLB offers substantial advantages as a disease control.

Our study is one of the first to use whole transcriptome to study transcriptional disruptions in LOAD. While Twine et al. examined transcription in LOAD using RNA-seq, our utilization of total RNA permitted us to observe changes in transcription in both protein coding and non-coding RNAs [8]. The ability of RNA-seq to detect novel splice junctions enabled us to discover that LOAD has a significant amount of alternative splicing across the genome and permits us to determine which processes were greatly affected by changes in splicing.

The samples used in this study allow us to identify transcriptional and epigenetic changes prior to the invasion of NFTs. In addition to the temporal pole tissue being a region where extracellular deposition of Aβ plaques (LOAD) and Lewy bodies (DLB) coincide, its use also enabled us to observe transcriptional changes prior to major visible destruction by NFTs. The temporal pole region is affected during the mid-stages (Braak stages 3 and 4) of LOAD and was collected, from patients that died during the early to mid-stage of pathological disease progression. In addition to limiting the patients to those at a Braak stage of 3 or 4, we also limited our analysis to examining a male only cohort so as to minimize confounding factors in so far as possible.

The process of demyelination destroys the myelin sheath of the nerve, results in damage to nerve conduction, and is associated with cognitive decline. Demyelination naturally occurs during aging; however, it is significantly increased in LOAD patients to the extent that its progression has recently been suggested for preclinical AD diagnosis [49]. While the extracted autopsy tissue contained both grey and white matter, several studies which have examined grey and white matter separately have observed that genes involved in the myelination process are similarly disrupted in LOAD suggesting that our findings are not just a result of the mixture of the two tissues [10]. The progression of demyelination spreads along the same path as the Aβ plaque development [50]. Alterations of the myelination network in LOAD have been detected in two previous studies [33, 39]. In addition to observing the myelination network that was also detected in previous studies, our study suggests that this network is altered specifically in LOAD rather than as part of the general neurodegenerative process. Analysis revealed that *ST18* (suppression of tumorigenicity 18) and *FRMD4B* (FERM domain containing 4B) were prominent hub genes within the myelination network. These two genes were also hubs in the myelination networks observed in two independent studies examining altered transcription in AD [33, 39]. *ST18* is a transcription factor that regulates neuronal differentiation. A genome-wide association study has also found *ST18* to be associated with cognitive decline in LOAD [51, 52]. Interestingly, we identified an anti-sense RNA (*ENSG00000253551.1*) with complementarity to the intronic region of *ST18* that, although after FDR it did not reach significance, is increased in LOAD ($p < 0.05$) when compared to both controls (Supplementary Table 7). Anti-sense RNAs can regulate genes with complementarity by binding to shared regions [53]. This anti-sense was positively correlated with *ST18* transcription($r = 0.93$; spearman's test), suggesting altered regulation of *ST18* in LOAD. In summary, our findings in combination with two previous independent studies suggest the myelination network changes are specific to LOAD.

Examination of individual coding and non-coding genes also led us to find LOAD specific differences in five genes involved in the myelination process. The *C10orf105, RARRES3,* and *ENSG0000249343.1* genes are all located in the myelination network detected above. While the functions of *C10orf105* and *ENSG0000249343.1* are unknown, *RARRES3* is known to increase in response to inflammation. *RARRES3* is also increased in obese patients with type

II diabetes, a disease that is a risk factor for LOAD [54]. While not directly in the myelination network, the functions of *WIF1* and *DIO2* are also related to myelination. DIO2 activates the thyroid hormone by converting the prohormone thyroxine (T4) to bioactive 3, 3', 5-triiodothyronine (T3) [37]. An increase in thyroxine "activates" myelination [55]. Similarly, a decrease in *WIF1* results in increased WNT-pathway activity an "essential driver" in myelinogenesis pathways [56]. Additionally, a recent GWAS has associated *WIF1* with hippocampal volume. Reduced hippocampal size is associated with AD [57] and this could, in part, be due to the decrease in the myelination facilitated by the decrease in *WIF1* transcription. Combined, the network analysis and the individual gene analysis suggest that disruption of the myelination process is LOAD specific.

Overall hypomethylation was observed across the genome in LOAD when compared to both controls. This finding is consistent with previous studies [58, 59]. In addition to transcription primarily being altered in the myelination network, significant changes of methylation also occurred within the network. Interestingly, the ratio of differentially methylated CpG sites to genes is considerably higher in the myelination network when compared to other genes. This hypomethylation corresponds to increased transcription observed in the myelination network. Studies have shown that DNA methylation in the gene body has a positive correlation with alternate splicing [60]. In the networks, we observed a statistically significant increase in the number of differentially methylated CpG sites in the gene body within the network (53%) compared to those in genes outside the network (47%) (χ^2; $p < 1.0e$-12) and we also observed a decrease in splicing in the network. These findings suggest an overall decrease in transcriptional diversity. While there is an increased amount of altered methylation within the myelination network, we found that the effect of methylation on transcription is weaker in LOAD when compared to both controls in both the promoter and the gene body. These findings suggest that could be pointing out slight alterations in the effects of DNA methylation within LOAD.

Using RNA-seq allowed us to also examine the role of ncRNAs within LOAD. When examining all genes, including protein coding and ncRNAs, for differential transcription, the most significant difference between LOAD and both controls was the ncRNA *ENSG0000249343.1*. Half of all genes with altered transcription in LOAD when compared to DLB are ncRNAs (Supplementary Fig. 3). Notably, both the

ST18 gene and *ST18-AS* were increased in LOAD when compared to normal controls; however, while *ST18* was slightly upregulated in LOAD when compared to DLB, it was *ST18-AS* that had the most significant increase in anti-sense RNA transcription between LOAD and the disease control. These findings suggest that noncoding RNAs play a significant role in helping to distinguish dementias and should be a focus for future study.

In summary, our approach using disease controls demonstrates that LOAD specific disruptions in transcription converge on genes and networks involved in myelination. Integrative analysis has also identified disruptions in DNA methylation within the myelination network. Interestingly, the networks and genes altered in LOAD are prominently involved in processes that are implicated in the progression of AD. The use of additional, different disease controls could allow the further honing in on genes specifically altered in LOAD and offers a very promising avenue for further investigations.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-141989.

REFERENCES

[1] Alzheimer's, Association (2012) 2012 Alzheimer's disease facts and figures. *Alzheimers Dement* **8**, 131-168.

[2] Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, James L et al. (1991) Segregation of a missense mutation in the amyloid precursor gene with familial Alzheimer's disease. *Nature* **349**, 704-706.

[3] Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K et al. (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973-977.

[4] Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K, Tsuda T, Mar

L, Foncin JF, Bruni AC, Montesi MP, Sorbi S, Rainero I, Pinessi L, Nee L, Chumakov I, Pollen D, Brookes A, Sanseau P, Polinsky RJ, Wasco W, Da Silva HA, Haines JL, Perkicak-Vance MA, Tanzi RE, Roses AD, Fraser PE, Rommens JM, St George-Hyslop PH (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754-760.

[5]   Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, Chi H, Lin C, Holman K, Tsuda T et al. (1995) Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* **376**, 775-778.

[6]   Tanzi RE (2012) The genetics of Alzheimer disease. *Cold Spring Harb Perspect Med* **2**, pii: a006296.

[7]   Williams C, Mehrian Shai R, Wu Y, Hsu YH, Sitzer T, Spann B, McCleary C, Mo Y, Miller CA (2009) Transcriptome analysis of synaptoneurosomes identifies neuroplasticity genes overexpressed in incipient Alzheimer's disease. *PLoS One* **4**, e4936.

[8]   Twine NA, Janitz K, Wilkins MR, Janitz M (2011) Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS One* **6**, e16266.

[9]   Colangelo V, Schurr J, Ball MJ, Pelaez RP, Bazan NG, Lukiw WJ (2002) Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: Transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *J Neurosci Res* **70**, 462-473.

[10] Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR, Landfield PW (2004) Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc Natl Acad Sci U S A* **101**, 2173-2178.

[11] Miller JA, Horvath S, Geschwind DH (2010) Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc Natl Acad Sci U S A* **107**, 12698-12703.

[12] Simpson JE, Ince PG, Shaw PJ, Heath PR, Raman R, Garwood CJ, Gelsthorpe C, Baxter L, Forster G, Matthews FE, Brayne C, Wharton SB (2011) Microarray analysis of the astrocyte transcriptome in the aging brain: Relationship to Alzheimer's pathology and APOE genotype. *Neurobiol Aging* **32**, 1795-1807.

[13] Braak H, Braak E (1991) Alzheimer's disease affects limbic nuclei of the thalamus. *Acta Neuropathol* **81**, 261-268.

[14] Alzheimer A, Stelzmann RA, Schnitzlein HN, Murtagh FR (1995) An English translation of Alzheimer's 1907 paper, "Uber eine eigenartige Erkankung der Hirnrinde". *Clin Anat* **8**, 429-431.

[15] American Psychiatric Association (1994) *Task Force on DSM-IV*, Diagnostic and Statistical Manual of Mental Disorders: DSM-IV. American Psychiatric Association, Washington, DC.

[16] Haroutunian V, Katsel P, Schmeidler J (2009) Transcriptional vulnerability of brain regions in Alzheimer's disease and dementia. *Neurobiol Aging* **30**, 561-573.

[17] McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, Cummings J, Duda JE, Lippa C, Perry EK, Aarsland D, Arai H, Ballard CG, Boeve B, Burn DJ, Costa D, Del Ser T, Dubois B, Galasko D, Gauthier S, Goetz CG, Gomez-Tortosa E, Halliday G, Hansen LA, Hardy J, Iwatsubo T, Kalaria RN, Kaufer D, Kenny RA, Korczyn A, Kosaka K, Lee VM, Lees A, Litvan I, Londos E, Lopez OL, Minoshima S, Mizuno Y, Molina JA, Mukaetova-Ladinska EB, Pasquier F, Perry RH, Schulz JB, Trojanowski JQ, Yamada M, Consortium on DLB

(2005) Diagnosis and management of dementia with Lewy bodies: Third report of the DLB Consortium. *Neurology* **65**, 1863-1872.

[18] Fujishiro H, Ferman TJ, Boeve BF, Smith GE, Graff-Radford NR, Uitti RJ, Wszolek ZK, Knopman DS, Petersen RC, Parisi JE, Dickson DW (2008) Validation of the neuropathologic criteria of the third consortium for dementia with Lewy bodies for prospectively diagnosed cases. *J Neuropathol Exp Neurol* **67**, 649-656.

[19] Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873-881.

[20] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

[21] Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**, R106.

[22] Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008-2017.

[23] Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S (2014) ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics* **30**, 428-430.

[24] Gentleman R. (2005) *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer Science+Business Media, New York.

[25] Langfelder P, Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatic* **9**, 559.

[26] Kadarmideen HN, Watson-haigh NS (2012) Building gene co-expression networks using transcriptomics data for systems biology investigations: Comparison of methods using microarray data. *Bioinformation* **8**, 855.

[27] Adamis D, Sahu S, Treloar A (2005) The utility of EEG in dementia: A clinical perspective. *Int J Geriatr Psychiatry* **20**, 1038-1045.

[28] Huang DW, Sherman BT, Lempicki RA (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protocols* **4**, 44-57.

[29] Diedenhofen B. (2013) cocor: Comparing correlations. http://r.birkdiedenhofen.de/pckg/cocor/.

[30] Dunn OJ, Clark V (1969) Correlation coefficients measured on the same individuals. *J Am Stat Assoc* **64**, 366-377.

[31] Webster JA, Gibbs JR, Clarke J, Ray M, Zhang W, Holmans P, Rohrer K, Zhao A, Marlowe L, Kaleem M, McCorquodale DS, 3rd, Cuello C, Leung D, Bryden L, Nath P, Zismann VL, Joshipura K, Huentelman MJ, Hu-Lince D, Coon KD, Craig DW, Pearson JV, Heward CB, Reiman EM, Stephan D, Hardy J, Myers AJ (2009) Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet* **84**, 445-458.

[32] Venables WN, Ripley BD (2002) *Modern applied statistics with S, 4th edition*, Springer, New York.

[33] Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezhnikov AA, Zhang C, Xie T, Tran L, Dobrin R (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707-720.

[34] Pollard KS, Gilbert HS, Ge Y, Taylor S, Dudoit S. multtest: Resampling-based multiple hypothesis testing. R package version 2.18.0

[35] Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.

[36] Bis JC, DeCarli C, Smith AV, van der Lijn F, Crivello F, Fornage M, Debette S, Shulman JM, Schmidt H, Srikanth V (2012) Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* **44**, 545.

[37] Buettner C, Harney JW, Larsen PR (2000) The role of seleno-cysteine 133 in catalysis by the human type 2 iodothyronine deiodinase. *Endocrinology* **141**, 4606-4612.

[38] Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, Laurance MF, Zhao W, Qi S, Chen Z, Lee Y, Scheck AC, Liau LM, Wu H, Geschwind DH, Febbo PG, Kornblum HI, Cloughesy TF, Nelson SF, Mischel PS (2006) Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci U S A* **103**, 17402-17407.

[39] Miller JA, Oldham MC, Geschwind DH (2008) A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci* **28**, 1410-1420.

[40] Flores K, Wolschin F, Corneveaux JJ, Allen AN, Huentelman MJ, Amdam GV (2012) Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics* **13**, 480.

[41] Humphries C, Kohli MA (2014) Rare Variants and Transcriptomics in Alzheimer disease. *Curr Genet Med Rep* **2**, 75-84.

[42] Gaig C, Valldeoriola F, Gelpi E, Ezquerra M, Llufriu S, Buongiorno M, Rey MJ, Marti MJ, Graus F, Tolosa E (2011) Rapidly progressive diffuse Lewy body disease. *Mov Disord* **26**, 1316-1323.

[43] Markesbery WR (1997) Neuropathological criteria for the diagnosis of Alzheimer's disease. *Neurobiol Aging* **18**, S13-S19.

[44] Daroff RB, Bradley WG (2012) *Bradley's Neurology in Clinical Practice*, 6th edition. Daroff RB, Fenichel GM, Jankovic J, Maziotta J, eds. Elsevier/Saunders, Philadelphia, PA.

[45] Iseki E (2004) Dementia with Lewy bodies: Reclassification of pathological subtypes and boundary with Parkinson's disease or Alzheimer's disease. *Neuropathology* **24**, 72-78.

[46] Tam CW, Burton EJ, McKeith IG, Burn DJ, O'Brien JT (2005) Temporal lobe atrophy on MRI in Parkinson disease with dementia: A comparison with Alzheimer disease and dementia with Lewy bodies. *Neurology* **64**, 861-865.

[47] Factor SA, Weiner WJ (2008) *Parkinson's Disease: Diagnosis and Clinical Management, 2nd*, Demos, New York.

[48] Spillantini MG, Crowther RA, Jakes R, Hasegawa M, Goedert M (1998) alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proc Natl Acad Sci U S A* **95**, 6469-6473.

[49] Carmeli C, Donati A, Antille V, Viceic D, Ghika J, von Gunten A, Clarke S, Meuli R, Frackowiak RS, Knyazeva MG (2013) Demyelination in mild cognitive impairment suggests progression path to Alzheimer's disease. *PLoS One* **8**, e72759.

[50] Bartzokis G (2004) Age-related myelin breakdown: A developmental model of cognitive decline and Alzheimer's disease. *Neurobiol Aging* **25**, 5-18; author reply 49-62.

[51] Kameyama T, Matsushita F, Kadokawa Y, Marunouchi T (2011) Myt/NZF family transcription factors regulate neuronal differentiation of P19 cells. *Neurosci Lett* **497**, 74-79.

[52] Yang J, Siqueira MF, Behl Y, Alikhani M, Graves DT (2008) The transcription factor ST18 regulates proapoptotic and proinflammatory gene expression in fibroblasts. *FASEB J* **22**, 3956-3967.

[53] Johnsson P, Ackley A, Vidarsdottir L, Lui WO, Corcoran M, Grander D, Morris KV (2013) A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat Struct Mol Biol* **20**, 440-446.

[54] Luchsinger JA, Tang MX, Stern Y, Shea S, Mayeux R (2001) Diabetes mellitus and risk of Alzheimer's disease and dementia with stroke in a multiethnic cohort. *Am J Epidemiol* **154**, 635-641.

[55] Calza L, Fernandez M, Giuliani A, Aloe L, Giardino L (2002) Thyroid hormone activates oligodendrocyte precursors and increases a myelin-forming protein and NGF content in the spinal cord during experimental allergic encephalomyelitis. *Proc Natl Acad Sci U S A* **99**, 3258-3263.

[56] Tawk M, Makoukji J, Belle M, Fonte C, Trousson A, Hawkins T, Li H, Ghandour S, Schumacher M, Massaad C (2011) Wnt/beta-catenin signaling is an essential and direct driver of myelin gene expression and myelinogenesis. *J Neurosci* **31**, 3729-3742.

[57] Frisoni GB, Ganzola R, Canu E, Rub U, Pizzini FB, Alessandrini F, Zoccatelli G, Beltramello A, Caltagirone C, Thompson PM (2008) Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. *Brain* **131**, 3266-3276.

[58] Chouliaras L, Mastroeni D, Delvaux E, Grover A, Kenis G, Hof PR, Steinbusch HW, Coleman PD, Rutten BP, van den Hove DL (2013) Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol Aging* **34**, 2091-2099.

[59] Condliffe D, Wong A, Troakes C, Proitsi P, Patel Y, Chouliaras L, Fernandes C, Cooper J, Lovestone S, Schalkwyk L, Mill J, Lunnon K (2014) Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain. *Neurobiol Aging* **35**, 1850-1854.

[60] Maunakea AK, Chepelev I, Cui K, Zhao K (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res* **23**, 1256-1269.