ORIGINAL ARTICLE

# Genes associated with the progression of neurofibrillary tangles in Alzheimer's disease

A Miyashita[1,10], H Hatsuta[2,10], M Kikuchi[3,10], A Nakaya[4], Y Saito[5], T Tsukie[3], N Hara[1], S Ogishima[6], N Kitamura[7], K Akazawa[7], A Kakita[8], H Takahashi[8], S Murayama[2], Y Ihara[9], T Ikeuchi[1], R Kuwano[1] and Japanese Alzheimer's Disease Neuroimaging Initiative

The spreading of neurofibrillary tangles (NFTs), intraneuronal aggregates of highly phosphorylated microtubule-associated protein tau, across the human brain is correlated with the cognitive severity of Alzheimer's disease (AD). To identify genes relevant to NFT expansion defined by the Braak stage, we conducted whole-genome exon array analysis with an exploratory sample set consisting of 213 human post-mortem brain tissue specimens from the entorinal, temporal and frontal cortices of 71 brain-donor subjects: Braak NFT stages 0 ($N = 13$), I–II ($N = 20$), III–IV ($N = 19$) and V–VI ($N = 19$). We identified eight genes, *RELN*, *PTGS2*, *MYO5C*, *TRIL*, *DCHS2*, *GRB14*, *NPAS4* and *PHYHD1*, associated with the Braak stage. The expression levels of three genes, *PHYHD1*, *MYO5C* and *GRB14*, exhibited reproducible association on real-time quantitative PCR analysis. In another sample set, including control subjects ($N = 30$), and in patients with late-onset AD ($N = 37$), dementia with Lewy bodies ($N = 17$) and Parkinson disease ($N = 36$), the expression levels of two genes, *PHYHD1* and *MYO5C*, were obviously associated with late-onset AD. Protein–protein interaction network analysis with a public database revealed that PHYHD1 interacts with MYO5C via POT1, and PHYHD1 directly interacts with amyloid beta-peptide 42. It is thus likely that functional failure of PHYHD1 and MYO5C could lead to AD development.

*Translational Psychiatry* (2014) **4**, e396; doi:10.1038/tp.2014.35; published online 10 June 2014

## INTRODUCTION

Alzheimer's disease (AD), the most common cause of dementia in the elderly, is a neurodegenerative disorder clinically characterized by progressive, insidious and irreversible cognitive decline. Major neuropathological features of AD brains are senile plaques and neurofibrillary tangles (NFTs). Senile plaques comprise the extracellular deposition of amyloid beta-peptide (Aβ), and NFTs the intracellular aggregation of highly phosphorylated microtubule-associated protein tau. These two lesions are thought to eventually lead to functional failure of synapses, synaptic loss and neuronal death in the AD brain. It has been demonstrated that the spreading of NFTs across the brain exhibits high correlation with the cognitive impairment status in AD.[1,2] The progression of NFTs is classified into six stages, I, II, III, IV, V and VI, based on the spatial distribution pattern of tangle-bearing neurons in the brain, which is known as the Braak NFT stage.[1,2] These stages could be further grouped as: the transentorhinal (I and II), limbic (III and IV) and neocortical (V–VI) stage groups, corresponding to normal cognition, cognitive impairment and dementia, respectively.[1,2]

Using the Braak NFT stage as an objective index of the neuropathological progression of AD, genome-wide gene expression studies involving post-mortem brain tissues have been carried out to discover the genes and molecular pathways related to the pathogenesis of AD.[3–7] A large number of genes were discovered to be AD-related ones through these studies. Bossers *et al.*[6] for the first time examined the gene expression pattern throughout the entire course of AD in the prefrontal cortex. In that study they demonstrated that the most prominent change in gene expression occurs between Braak NFT stages II and III, and genes related to synaptic activity are primarily affected around these stages.[6]

We considered that the altered gene expression associated with NFT expansion may be useful for monitoring AD progression. Thus, we here conducted whole-genome gene expression analysis using 213 human post-mortem brain tissue specimens from three brain regions (BRs), the entorhinal (EC), temporal (TC) and frontal (FC) cortices, of 71 Japanese brain-donor subjects[8,9] to identify the genes associated with the spreading of NFTs in each BR. The disease specificity of the genes identified was ascertained by using healthy control subjects ($N = 30$) and patients with late-onset AD (LOAD, $N = 37$), dementia with Lewy bodies (DLB, $N = 17$) and Parkinson disease (PD, $N = 36$). Next, protein–protein interaction (PPI) network analysis was carried out to determine whether or not physical connections exist among the proteins encoded by the NFT-associated genes identified.

## MATERIALS AND METHODS

### Post-mortem brain tissues

This study was approved by the Institutional Review Board of Niigata University and by all participating institutes. All subjects were anonymously analyzed.

[1]Department of Molecular Genetics, Bioresource Science Branch, Center for Bioresources, Brain Research Institute, Niigata University, Niigata, Japan; [2]Department of Neuropathology, Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan; [3]Research Association for Biotechnology, Tokyo, Japan; [4]Center for Transdisciplinary Research, Niigata University, Niigata, Japan; [5]Department of Pathology, National Center Hospital of Neurology and Psychiatry, Tokyo, Japan; [6]Department of Health Record Informatics, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan; [7]Department of Medical Informatics, Niigata University, Niigata, Japan; [8]Department of Pathology, Brain Research Institute, Niigata University, Niigata, Japan and [9]Department of Neuropathology, Faculty of Life and Medical Sciences, Doshisha University, Kyoto, Japan. Correspondence: Dr A Miyashita or Professor R Kuwano, Department of Molecular Genetics, Brain Research Institute, Niigata University, 1-757 Asahimachi, Niigata 951-8585, Japan.
E-mails: miyashi@bri.niigata-u.ac.jp or ryosun@bri.niigata-u.ac.jp
[10]These authors contributed equally to this work.
Received 14 January 2014; revised 27 March 2014; accepted 22 April 2014

**Table 1.** Genes showing both $P_{FDR} < 5.00E-02$ and fold change $> |1.5|$ in Com-NFT(BR)

| Comparison ID | Gene | $P_{FDR}$ | Fold change | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | I–II vs $0_{ref}$ | III-IV vs $0_{ref}$ | V–VI vs $0_{ref}$ | III-IV vs I-II$_{ref}$ | V–VI vs I-II$_{ref}$ | V–VI vs III-IV$_{ref}$ | I–VI vs $0_{ref}$ | III-VI vs 0-II$_{ref}$ | V–VI vs 0-IV$_{ref}$ |
| Com-NFT(E) | *RELN* | **2.96E−09** | -1.26 | −1.09 | −**1.99** | 1.16 | −**1.58** | −**1.82** | −1.40 | −1.31 | −**1.79** |
| | *PTGS2* | **3.88E−05** | −1.36 | −**1.62** | −**1.75** | −1.19 | −1.28 | −1.08 | −**1.57** | −1.44 | −1.34 |
| | *MYO5C* | **3.67E−04** | 1.34 | 1.38 | **1.84** | 1.03 | 1.38 | 1.33 | **1.50** | 1.38 | **1.50** |
| | *TRIL* | **3.64E−03** | 1.48 | **1.53** | **1.75** | 1.03 | 1.18 | 1.15 | **1.58** | 1.34 | 1.33 |
| | *DCHS2* | **2.43E−02** | **1.60** | 1.41 | 1.31 | −1.13 | −1.23 | −1.08 | 1.44 | 1.07 | −1.01 |
| | *GRB14* | **2.78E−02** | −1.14 | 1.13 | −1.48 | 1.29 | −1.29 | −**1.67** | −1.14 | −1.07 | −1.47 |
| | *NPAS4* | **2.96E−02** | −**1.91** | −**1.81** | −**1.98** | 1.06 | −1.03 | −1.09 | −**1.90** | −1.37 | −1.31 |
| Com-NFT(T) | *PHYHD1* | **3.18E−03** | 1.39 | 1.35 | **1.59** | −1.03 | 1.14 | 1.17 | 1.44 | 1.24 | 1.28 |
| Com-NFT(F) | *NPAS4* | **3.09E−04** | −**2.73** | −**2.89** | −**2.43** | −1.06 | 1.12 | 1.19 | −**2.68** | −**1.61** | −1.22 |

Abbreviation: ref, reference. $P_{FDR}$ calculated by alt-splicing ANOVA is shown in underlined boldface type. Fold change values exhibiting $>|1.5|$ are depicted in boldface type.

We prepared an exploratory sample set, ROW, comprising 71 Japanese brain-donor subjects (Supplementary Table S1). The neuropathological diagnoses of all these subjects are given in Supplementary Table S8. Based on the NFT stage advocated by Braak H and Braak E,[1] the subjects were divided into four groups: Braak NFT stages 0 ($N = 13$), I–II ($N = 20$), III–IV ($N = 19$) and V–VI ($N = 19$). Tissues dissected from three BRs, EC, TC and FC, were used. In total, 213 brain tissue specimens ($= 71$ subjects $\times 3$ BRs) were eventually involved in this study.

To determine whether expression variations of genes identified in the ROW set were disease-specific or not, we constructed another sample set, NP, comprising 120 well-defined brain-donor subjects (Supplementary Table S7). The NP set was composed of control subjects ($N = 30$), and patients with LOAD ($N = 37$), DLB ($N = 17$) and PD ($N = 36$). Brain specimens from FC were utilized in this sample set.

## Genomic DNA extraction and *APOE* genotyping
Preparation of genomic DNA from FC and determination of *APOE* genotypes were described in our previous paper.[10]

## Total RNA extraction and quality control
Total RNA from brain tissues was extracted with a TRIzol Plus RNA Purification System (Life Technologies, Carlsbad, CA, USA). Genomic DNA was removed through on-column DNase I treatment during the RNA preparation. For determination of the RNA integrity number (1 (totally degraded) to 10 (intact)),[11] a 2100 Bioanalyzer instrument was used with the RNA 6000 Pico Assay (Agilent, Santa Clara, CA, USA). We fluorometrically determined the concentration of total RNA with a Quant-iT RiboGreen RNA Assay Kit (Life Technologies).

## Whole-genome gene expression profiling
For a genome-wide survey of transcripts associated with Braak NFT stages and BRs, we used GeneChip Human Exon 1.0 ST Arrays (Affymetrix, Santa Clara, CA, USA). All reactions were carried out according to the manufacturer's instruction manual (P/N 701880, Rev. 4).

The initial raw data (DAT files) were processed into CEL files via the Affymetrix GeneChip Operating Software. The CEL files were imported into Partek Genomics Suite 6.4 (Partek, St Louis, MO, USA) and then normalized, and their backgrounds were corrected by means of a robust multi-array average method[12] with adjustment for GC content. The expression levels of all probesets in the CEL files were $\log_2$-transformed. Core meta-probesets including 232,479 probesets (Affymetrix reference file name: HuEx-1_0-st-v2.r2.dtl.hg18.core.mps) were used. Among them, we excluded probesets that did not have an official gene symbol and did not exhibit a maximum signal intensity of 5.0 across all exon arrays ($N = 213$). Consequently, 190,447 probesets per CEL file were analyzed in further examinations.

We compared the gene expression levels among the four Braak NFT stage groups in each BR, designated as Comparison(Com)-NFT(BR): Com-NFT(E) in EC, Com-NFT(T) in TC and Com-NFT(F) in FC (Supplementary Figure 1). Also, in the four Braak NFT stage groups, the gene expression levels were compared among the three BR groups, referred to as Com-BR(NFT): Com-BR(0) in Braak NFT stage 0, Com-BR(I-II) in Braak NFT stages I–II, Com-BR(III–IV) in Braak NFT stages III–IV and Com-BR(V–VI) in Braak NFT stages V–VI (Supplementary Figure 1). For each gene, alternative splicing analysis of variance (alt-splicing ANOVA) P-values (P) were computed. A method to control the false discovery rate (FDR)[13] was applied for correct multiple testing ($P_{FDR}$): alt-splicing ANOVA $P_{FDR}$ of 5.00E−02 was set as the significant threshold (Supplementary Table S9). Additionally, we calculated nine items in Com-NFT(BR) (Table 1) and three items in Com-BR(NFT) (Supplementary Tables S3 and S4) as to the fold changes. These analyses were implemented into Partek Genomics Suite 6.4 (Partek). We selected genes exhibiting both $P_{FDR} < 5.00E-02$ and fold change $\geqslant |1.5|$ for real-time quantitative PCR (qRT-PCR) analysis with TaqMan assays, as described below (Supplementary Figure 1).

## RT–qPCR
Samples were subjected to RT–qPCR amplification with a TaqMan Gene Expression Assay (Life Technologies) on an ABI PRISM 7900 HT instrument (Applied Biosystems, Carlsbad, CA, USA). The details are given under Supplementary Methods.

## Gene ontology and pathway analyses
To systematically elucidate the biological functions of genes discovered in Com-BR(NFT), we performed gene ontology (GO) and pathway analyses with The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7.[14] GO and pathway terms showing unadjusted P-values $< 1.00E-01$ were selected. For pathway analysis, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database[15] was used to reveal physical and/or functional interactions among the genes.

## Hierarchical cluster analysis
To group genes based on similar gene expression patterns across both the four Braak NFT stage groups (0, I–II, III–IV and V–VI) and the three BRs (EC, TC and FC), hierarchical cluster analysis was conducted using Partek Genomics Suite 6.4. After calculation of the z-score for each gene, gene clustering was performed with an average linkage method with the Euclidean distance.

## PPI network analysis
Proteins encoded by genes identified via exon array analysis with the ROW set (Supplementary Table S1) were subjected to PPI network analysis. A comprehensive protein interaction database, BioGRID,[16] was referred to. In this database (build 3.2.104 (September 2013)), 17 841 unique physical protein interactions have been deposited and examined to construct PPI networks. We retrieved one-hop neighborhood proteins and connected them to each other as edges.
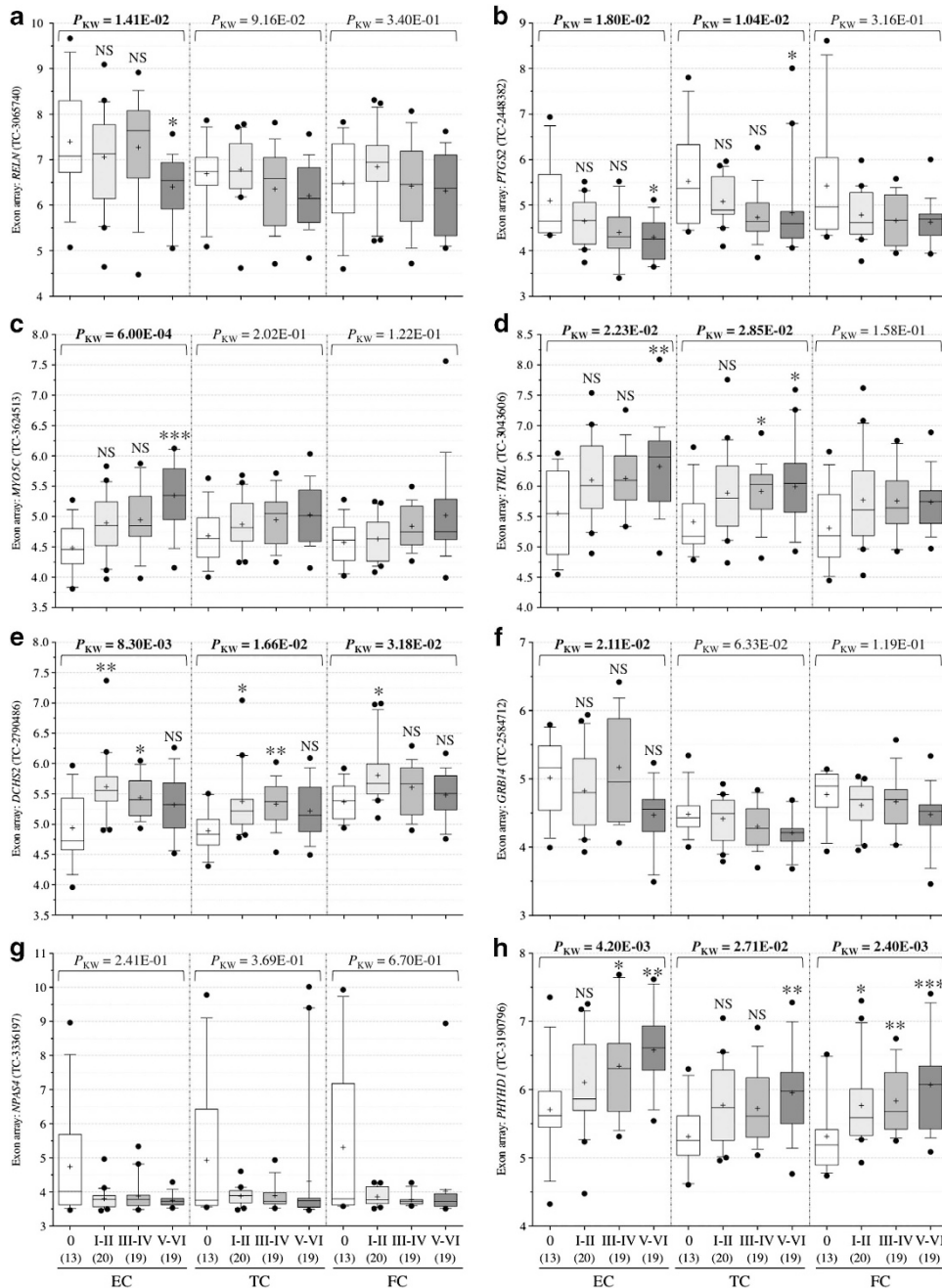
**Figure 1.** Comparison of the expression levels of eight genes, *RELN*, *PTGS2*, *MYO5C*, *TRIL*, *DCHS2*, *GRB14*, *NPAS4* and *PHYHD1*, among four Braak NFT stage groups (0, I–II, III–IV and V–VI) in three brain regions (BRs) (EC, TC and FC). For this analysis, we used data derived from genechips with the ROW set (Supplementary Tables S1 and S8). Box-and-whisker plots show the distributions of the gene-level expression levels (log$_2$-transformed values) of these genes in every Braak NFT stage group (0, I–II, III–IV and V–VI): upper horizontal line of box, 75th percentile; lower horizontal line of box, 25th percentile; horizontal bar within box, median; upper horizontal bar outside box, 90th percentile; lower horizontal bar outside box, 10th percentile; dot, outlier; plus, mean. For the four-group comparison (0 vs I–II vs III–IV vs V–VI) in each BR, the Kruskal–Wallis test was conducted. When statistical significance ($P_{KW} < 5.00E-02$, depicted in boldface type) was observed, Dunn's multiple comparison test was used as a *post hoc* test. Braak NFT stage 0 was used as a reference group for this test: multiplicity-adjusted *P*-values were calculated (*$P < 5.00E-02$, **$P < 1.00E-02$ and ***$P < 1.00E-03$). The numbers of subjects included in each Braak NFT stage group are shown in parentheses. (**a**) *RELN* (transcript cluster ID [TC]-3065740); (**b**) *PTGS2* (TC-2448382); (**c**) *MYO5C* (TC-3624513); (**d**) *TRIL* (TC-3043606); (**e**) *DCHS2* (TC-2790486); (**f**) *GRB14* (TC-2584712); (**g**) *NPAS4* (TC-3336197) and (**h**) *PHYHD1* (TC-3190796). NS, not significant; $P_{KW}$, *P*-value computed by means of the Kruskal–Wallis test.

## Statistical analysis

The gender (female vs male) and *APOE-ε4* (*ε4* allele carrier vs non-carrier) distributions were assessed by means of $\chi^2$ and Fisher's exact tests, respectively. For comparisons among three or four groups, we applied the Kruskal–Wallis test ($P_{KW}$), that is, Braak NFT stages (0 vs I–II vs III–IV vs V–VI) and BRs (EC vs TC vs FC). If statistical significance was observed in the test, Dunn's multiple comparison test was adopted as a *post hoc* test, and multiplicity adjusted *P* for each comparison was computed. By means of one-way ANOVA ($P_{ANOVA}$), differences of gene expression levels among four subject groups, control, LOAD, DLB and PD, were examined, followed

by Sidak's multiple-comparison *post hoc* test for each comparison. We carried out the Mann–Whitney test ($P_{MW}$) for two-group comparisons, that is, *APOE* (ε4 non-carrier vs carrier) and gender (female vs male). Using Pearson's coefficient value, *R*, we examined the correlation between two groups. Multiple linear regression analysis was performed to determine whether or not the expression levels of genes identified were significantly associated with Braak NFT stage (0, I–II, III–IV and V–VI), with adjustment for three covariates, BR (EC, TC and FC), gender (male and female) and AAD.

The statistical analyses described above were implemented into Prism version 6.0c for Mac OS X (GraphPad Software, La Jolla, CA, USA) and R software (version 3.0.3). We considered $P < 5.00E-02$ to be statistically significant.

## RESULTS

### Genes associated with NFT spreading

To identify differentially expressed genes as to Braak NFT stages, we conducted whole-genome gene expression analysis using exon arrays with the ROW set (Supplementary Tables S1 and S8). In the three BRs (EC, TC and FC), the gene expression levels were compared among the four Braak NFT stage groups, 0, I-II, III-IV and V-VI, by means of alt-splicing ANOVA (Supplementary Figure 1). We found that the expression levels of eight genes ($P_{FDR} < 5.00E-02$ and fold change $\geqslant |1.5|$) were altered as to Braak NFT stage: seven genes, *RELN* (reelin, 7q22), *PTGS2* (prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase), 1q25.2-q25.3), *MYO5C* (myosin VC, 15q21), *TRIL* (TLR4 interactor with leucine-rich repeats, 7p14.3), *DCHS2* (dachsous 2 (Drosophila), 4q31.3), *GRB14* (growth factor receptor-bound protein 14, 2q22–q24), and *NPAS4* (neuronal PAS domain protein 4, 11q13), in Com-NFT(E), one gene, *PHYHD1* (phytanoyl-CoA dioxygenase domain containing 1, 9q34.11), in Com-NFT(T), and one gene, *NPAS4*, in Com-NFT(F) (Table 1 and Supplementary Table S2). The expression patterns of the eight genes are presented as entire transcripts in Supplementary Figure 3.

By means of the Kruskal–Wallis test, the expression levels of the eight genes were compared among the four Braak NFT stage groups in each BR (Figure 1). Seven genes, *RELN*, *PTGS2*, *MYO5C*, *TRIL*, *DCHS2*, *GRB14* and *PHYHD1*, were significantly associated with Braak NFT stage in at least one BR (Figure 1). As to *DCHS2* (Figure 1e) and *PHYHD1* (Figure 1h), we observed significance in every BR.

We also examined whether or not the expression levels of the eight genes differed depending on the *APOE*-ε4 carrier status (ε4 non-carrier vs carrier) and gender (female vs male) in each BR. The difference in the gene expression level of *PHYHD1* was significant between ε4 non-carriers and carriers: $P_{MW} = 1.03E-02$ in TC and $P_{MW} = 1.87E-02$ in FC. There were no significant differences between female and male subjects for any of the eight genes.

In AD brains, the major histopathological hallmarks in addition to senile plaques and NFTs are neuronal loss and gliosis, which lead to population changes of brain cells.[17] Thus, we considered that the expression levels of neural cell-type-specific genes should be carefully checked. The genes listed below were examined: *AQP4*, *GFAP* and *S100B* for astrocytes, *AIF1*, *CD68*, *EMR1* and *LGALS3* for microglia, *MAG*, *MBP*, *MOG* and *SOX10* for oligodendrocytes, and *CHGA*, *ENOS2*, *NEFL*, *NEFM*, *NEFH*, *SNAP25* and *SYT1* for neurons. In the ROW set (Supplementary Table S1), none of these genes showed significant expression changes across Braak NFT stages (Supplementary Figure 2).

### Genes associated with BRs

Using the ROW set (Supplementary Tables S1 and S8), we also compared the gene expression levels among the three BRs, EC, TC and FC, in each Braak NFT stage (Supplementary Figure 1). A total of 357 genes exhibited alt-splicing ANOVA $P_{FDR} < 5.00E-02$ and fold change $\geqslant |1.5|$ (Supplementary Tables S2 and S4). The overlapping of these genes is represented as a Venn diagram in Supplementary Figure 5: 15 sections were generated, which were

individually designated as Sections (Sec) A – O. The expression levels of 36 genes in Sec-A (Supplementary Figure 5) were found not to be affected by the Braak NFT stage. On the other hand, it was found that in total 193 genes in Sec-L (91 genes), M (10 genes), N (25 genes) and O (67 genes) strongly depend on the Braak NFT stage (Supplementary Figure 5). For the genes in Sec-A, L, M, N and O, we performed GO and KEGG pathway analyses: the results are given in Supplementary Tables S5 and S6, respectively.

We paid particular attention to eight genes, *RELN*, *PTGS2*, *MYO5C*, *TRIL*, *DCHS2*, *GRB14*, *NPAS4* and *PHYHD1*, discovered in Com-NFT (BR) (Table 1, Figure 1 and Supplementary Figure 3). Thus, we checked alterations in their expression levels in Com-BR(NFT). We detected apparent decreases in the gene exprsssion levels of *GRB14* and *PHYHD1* in TC compared with those in EC: *GRB14*, $P_{FDR} = 9.40E-09$ and fold change $= -1.82$ in Com-BR(III–IV); *PHYHD1*, $P_{FDR} = 4.20E-02$ and fold change $= -1.53$ in Com-BR (III–IV); and $P_{FDR} = 7.94E-04$ and fold change $= -1.55$ in Com-BR (V-VI) (Supplementary Table S3). *GRB14* and *PHYHD1* were included in Sec-N and Sec-J, respectively (Supplementary Figure 5).

### Clustering of genes associated with Braak NFT stages according to their expression patterns

To categorize the eight genes according to their gene expression patterns, we performed hierarchical cluster analysis (Figure 2a). Four major clusters were observed: *RELN* and *GRB14* in Cluster-1, *NPAS4* and *PTGS2* in Cluster-2, *DCHS2* in Cluster-3, and *MYO5C*, *PHYHD1* and *TRIL1* in Cluster-4 (Figure 2a). The expression patterns of these genes across Braak NFT stages are presented in Figures 2b–i. In Cluster-1 including *RELN* (Figure 2b) and *GRB14* (Figure 2c), gradual decreases in the gene expression levels were observed with Braak NFT stage progression. *NPAS4* (Figure 2d) and *PTGS2* (Figure 2e) in Cluster-2 exhibited decreases in their expression levels along with the Braak NFT stage progression; in particular, the degree of the decrease from Braak NFT stages 0 to I–II was remarkable for *NPAS4* (Figure 2d). Transitory increases in the gene expression levels of *DCHS2* in Cluster-3 were detected from Braak NFT stage 0 to I–II, followed by gradual decreases in its expression toward later stages (Figure 2f). The gene expression levels of *MYO5C* (Figure 2g), *PHYHD1* (Figure 2h) and *TRIL* (Figure 2i) gradually increased along with the Braak NFT stage progression.

In the three BRs, we investigated the correlation of the expression levels of genes included in Clusters-1, 2 and 4 (Supplementary Figure 4). Only for the correlation between *MYO5C* and *TRIL* in Cluster-4 was significance not detected (Supplementary Figure 4j–l).

### PPI network including genes identified in the exploratory anlysis

We attempted to characterize physical PPI networks in which eight gene products, RELN, PTGS2, MYO5C, TRIL, DCHS2, GRB14, NPAS4 and PHYHD1, are relevant (Figure 3). The BioGRID database[16] was screened for this analysis. It was found that all eight proteins have at least one PPI. We found that (1) RELN, DCHS2 and PTGS2 interact via COPS5 (COP9 signalosome subunit 5); (2) the interaction between two proteins, MYO5C and PHYHD1, is mediated by POT1 (protection of telomeres 1); (3) PHYHD1 directly interacts with Aβ42,[18] which is denoted as 'APP' (amyloid beta (A4) precursor protein) in Figures 3; and (4) GRB14 interacts with APP via IGF1R (insulin-like growth factor 1 receptor).

### RT–qPCR analysis of genes identified in the exploratory anlysis

We attempted to replicate the association of eight genes (Table 1, Figure 1 and Supplementary Figure 3) identified in the ROW set (Supplementary Tables S1 and S8) with Braak NFT stage using the TaqMan method. We first assessed the correlation between exon array and TaqMan data. Significantly high positive correlations
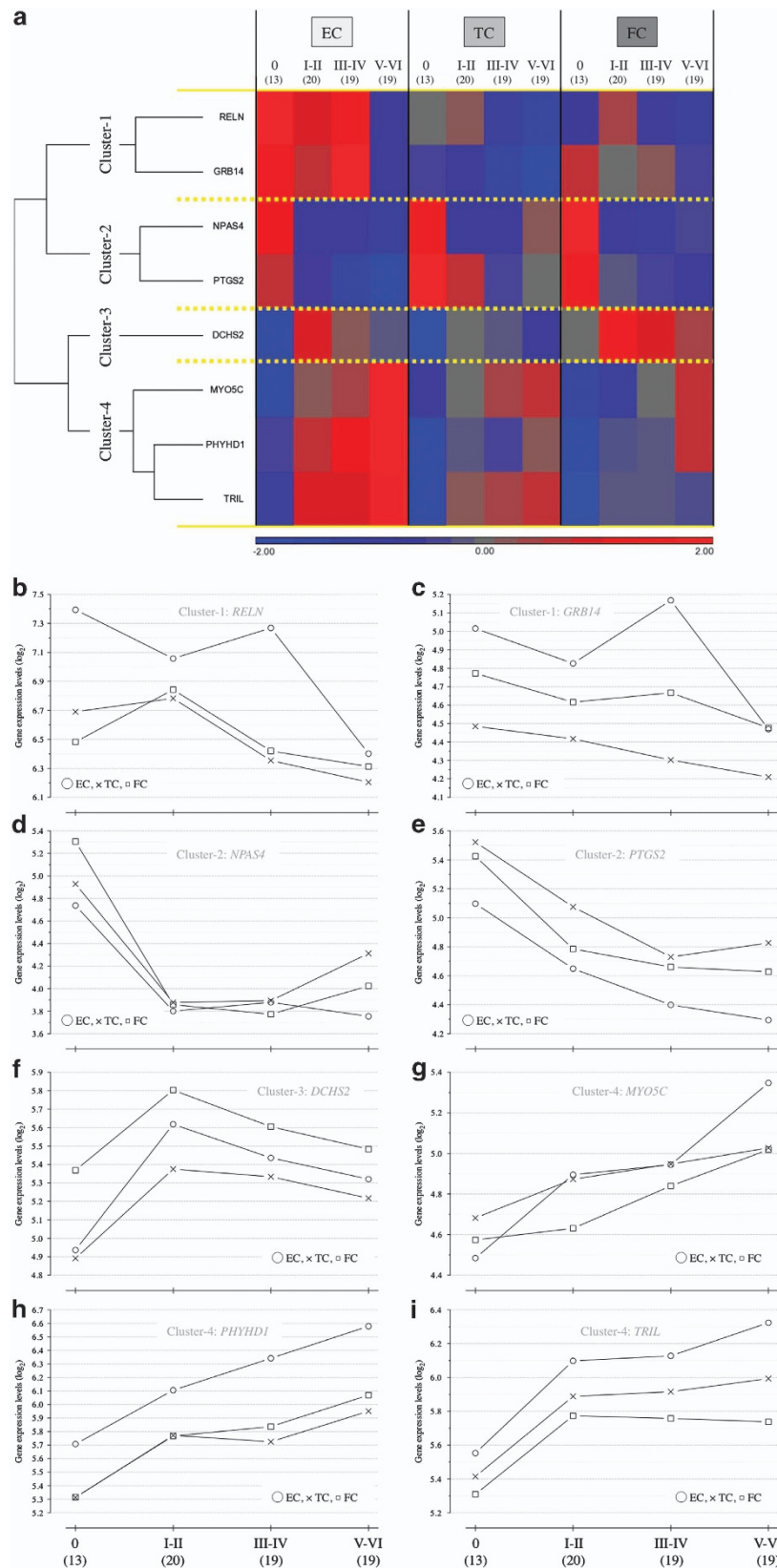
**Figure 2.** Hierarchical clustering of the eight genes, *RELN*, *PTGS2*, *MYO5C*, *TRIL*, *DCHS2*, *GRB14*, *NPAS4* and *PHYHD1*. (**a**) These genes were grouped in accordance with their *Z*-scored gene expression levels across Braak NFT stages. Red and blue indicate upregulated and downregulated gene expression levels, respectively. (**b–i**) Expression patterns of the eight genes across Braak NFT stages: (**b**), *RELN* (Cluster-1); (**c**) *GRB14* (Cluster-1); (**d**) *NPAS4* (Cluster-2); (**e**) *PTGS2* (Cluster-2); (**f**) *DCHS2* (Cluster-3); (**g**) *MYO5C* (Cluster-4); (**h**) *PHYHD1* (Cluster-4); and (**i**) *TRIL* (Cluster-4). The mean log$_2$-transformed gene expression levels of each Braak NFT stage group are plotted. The numbers of subjects included in each Braak NFT stage group are shown in parentheses.
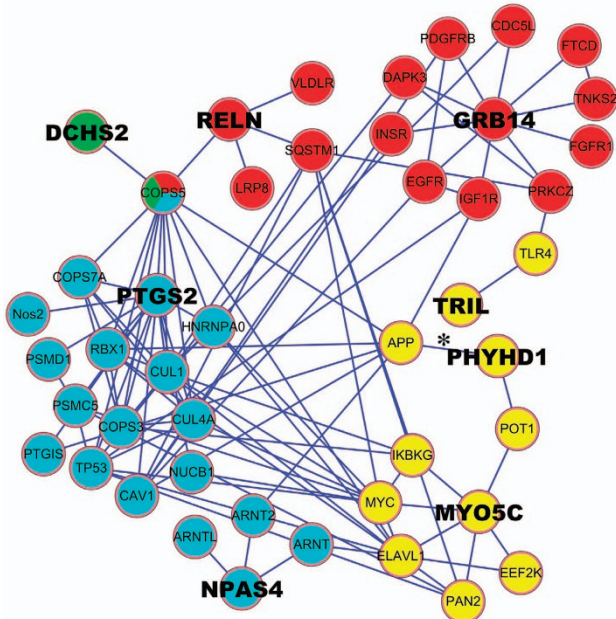
**Figure 3.** Protein–protein interaction network including the eight proteins encoded by the eight genes, *RELN*, *PTGS2*, *MYO5C*, *TRIL*, *DCHS2*, *GRB14*, *NPAS4* and *PHYHD1*. For construction of the network, we used a protein–protein interaction database, BioGrid (build 3.2.104 (September 2013)).[16] As TRIL was not included in build 3.2.104 of this database, we manually added TRIL to this network via TLR4, which directly interacts with TRIL.[36] The eight proteins, RELN, PTGS2, MYO5C, TRIL, DCHS2, GRB14, NPAS4 and PHYHD1, are highlighted in boldface type and are colored according to the clusters presented in Figure 2: Cluster-1 (RELN and GRB14), red; Cluster-2 (NPAS4 and PTGS2), turquoise; Cluster-3 (DCHS2), green; and Cluster-4 (PHYHD1, MYO5C and TRIL), yellow. Two proteins, COPS5 and POT1, were found to mediate interactions among multiple proteins identified in this study: COPS5, RELN, DCHS2 and PTGS2; and POT1, PHYHD1 and MYO5C. Note that APP, sequentially cleaved by β- and γ-secretases toward neurotoxic Aβ production, directly interacts with one of the proteins discovered in this study, PHYHD1. The edge between APP and PHYHD1 is indicated by an asterisk.

were observed for all eight genes ($P < 1.00E-04$): range of $R^2$, 0.25–0.85. Then, we performed comparisons among the four Braak NFT stage groups, 0, I–II, III–IV and V–VI, for each gene. The expression levels of three genes, *MYO5C* (in EC), *GRB14* (in EC) and *PHYHD1* (in TC), were confirmed to be obviously associated with Braak NFT stage (Supplementary Figure 6).

**Disease specificity of three genes, MYO5C, GRB14 and PHYHD1**

We investigated whether or not disease-dependent gene expression changes occur for three genes, *PHYHD1*, *MYO5C* and *GRB14*. The NP set (Supplementary Table S7) was used for this analysis. Global significant differences among the four disease groups (control, LOAD, DLB and PD) were detected for all three genes: *PHYHD1* (Figure 4a), $P_{ANOVA} = 8.60E-03$ (RPS17) and $1.88E-02$ (CASC3); *MYO5C* (Figure 4b), $P_{ANOVA} = 7.60E-03$ (RPS17) and $2.76E-02$ (CASC3); and *GRB14* (Figure 4c), $P_{ANOVA} = 4.71E-02$ (GUSB), $3.75E-02$ (RPS17) and $9.10E-03$ (CASC3). There were significant expression alterations between the control and LOAD groups, but not between the control and PD ones, in *PHYHD1* (Figure 4b) and *MYO5C* (Figure 4b). In patients with DLB in comparison with control subjects, we observed significantly increased gene expression of *PHYHD1* (RPS17 and CASC3) (Figure 4a). Concerning *GRB14*, significant gene expression changes were not detected in any comparisons between two
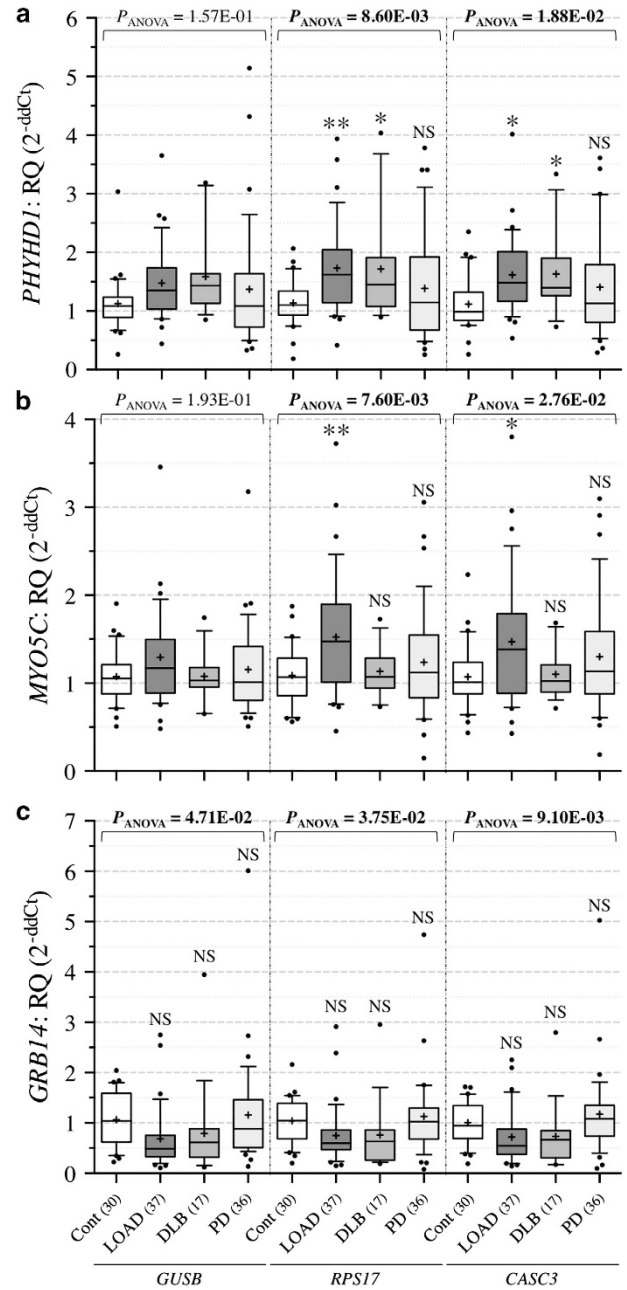


**Figure 4.** Disease specificity of the expression levels of three genes, *PHYHD1*, *MYO5C* and *GRB14*. RT–qPCR analysis of the three genes was performed in the NP set (Supplementary Table S7). As endogenous control genes for normalization of the target gene expression levels, we used GUSB, RPS17 and CASC3. Box-and-whisker plots show the distributions of the expression levels of *PHYHD1* (**a**), *MYO5C* (**b**) and *GRB14* (**c**) in every disease group: upper horizontal line of box, 75th percentile; lower horizontal line of box, 25th percentile; horizontal bar within box, median; upper horizontal bar outside box, 90th percentile; lower horizontal bar outside box, 10th percentile; dot, outlier; plus, mean. For four-group comparison for each endogenous control gene, one-way ANOVA was conducted. When statistical significance ($P_{ANOVA} < 5.00E-02$, depicted in boldface type) was observed, Sidak's multiple-comparison *post hoc* test was applied. Control subjects (Cont) were used as a reference group for this test: multiplicity-adjusted *P*-values were calculated ($*P < 5.00E-02$ and $**P < 1.00E-02$). The numbers of subjects included in each disease group are shown in parentheses. AD, Alzheimer's disease; DLB, dementia with Lewy bodies; PD, Parkinson disease; NS, not significant; RQ, relative quantity; ddCt, delta delta cycle threshold.

groups; however, the gene expression levels of LOAD and DLB patients were found to exhibit a decreasing tendency compared with that of control subjects (Figure 4c).

## DISCUSSION

NFTs are one of the neuropathological hallmarks characteristic of AD brains. Their spreading through the brain is highly correlated with the cognitive state of AD.[1,2] We thought that gene expression variation associated with NFT expansion may be useful for monitoring AD progression. Therefore, in this study we attempted to discover transcripts associated with Braak NFT stage.[1] Through whole-genome gene expression analysis, we identified eight genes, RELN, PTGS2, MYO5C, TRIL, DCHS2, GRB14, NPAS4 and PHYHD1, as NFT-associated ones (Table 1,Figure 1 and Supplementary Figure 3). Genes related to the phosphorylation and dephosphorylation of tau, such as GSK3B and PPP1CA,[19,20] did not show significant association with Braak NFT stage here. Concerning three genes, RELN[21,22], PTGS2[23,24] and DCHS2[25], there have been reports mentioning their involvment in AD. Three proteins, RELN, PTGS2 and DCHS2, directly interact with COPS5 (Figure 3), which directly interacts with APP and thus increases Aβ production.[26] However, to our knowledge, the remaining five genes, MYO5C, TRIL, GRB14, NPAS4 and PHYHD1, are novel to AD. Among these five genes, we were able to replicate the association of three genes, PHYHD1, MYO5C and GRB14, with Braak NFT stages by means of RT–qPCR analysis (Supplementary Figure 6). Multiple linear regression analysis with adjustment for BR, gender and AAD supported the association of the expression levels of the three genes with Braak NFT (Supplementary Table S10). In particular, the gene expression levels of PHYHD1 and MYO5C were found to evidently change in patients with LOAD compared with that in control subjects (Figure 4). It is likely that functional failure of PHYHD1 and MYO5C leads to AD development.

On PPI network analysis (Figure 3), we obtained evidence that PHYHD1 directly interacts with $A\beta_{42}$,[18] which is cleaved from APP through sequential proteolysis by β- and γ-secretases. PHYHD1 encodes three protein isoforms, PHYHD1A (291 amino acids (a.a.)), PHYHD1B (297 a.a.) and PHYHD1C (270 a.a.), generated through different splicing. Among them, PHYHD1A functions as a Fe(II) and 2-oxoglutarate (2OG)-dependent oxygenase,[27] which catalyzes the conversion of 2OG to succinate and $CO_2$.[28] The 2OG oxygenases comprise a large superfamily, consisting of $> 60$ family members, that is involved in various biochemical reactions including fatty acid metabolism via the phytanic acid α-oxidation pathway.[28] It is suggested, on the basis of biochemical and structural analyses, that PHYHD1A may interact with some fatty acids or their derivatives, and thus be involved in their hydroxylation.[27] Fatty acids, especially omega-3 and -6 fatty ones, have an influence on Aβ metabolism.[29] Altered gene expression of PHYHD1 is likely to affect the efficiency of metabolism of fatty acids, which might finally influence several neuronal activities in the human brain.

MYO5C was found to interact with PHYHD1 via POT1 (Figure 3). There were significant correlations between the gene expression levels of MYO5C and PHYHD1 (Supplementary Figure 4g–i). The patterns of gene expression variations across the four Braak NFT stages were almost the same for both genes (Figures 2a, g and h). MYO5C is the third member of the class V myosins (actin-based motor proteins), and exhibits ~ 50% similarity in its overall a.a. sequence to those of MYO5A and MYO5B.[30] MYO5C[30,31] is suggested to have an important role in secretory granule trafficking involving transferrin, a well-known protein involved in iron homeostasis.[32] RAB10, one of the RAB proteins influencing vesicle trafficking pathways, interacts with MYO5C as well as MYO5A and MYO5B.[33] Functional failure of proteins related to intracellular trafficking, such as SORL1, is thought to lead to AD development along with abnormal Aβ production and deposition.[34,35] Although at present there have been no reports

mentioning that MYO5C is involved in AD pathogenesis, our findings indicate that MYO5C could be a promising candidate for AD.

In conculsion, eight genes, RELN, PTGS2, MYO5C, TRIL, DCHS2, GRB14, NPAS4 and PHYHD1, were identified as NFT-associated ones on whole-genome gene expression analysis. PHYHD1 and MYO5C were replicated as associated genes with not only Braak NFT stage but also LOAD beside other neurodegenerative diseases, DLB and PD. These two genes have attractive biological functions, of which failure may lead to AD development. Further functional examinations of PHYHD1 and MYO5C will provide insights into the mechanisms underlying AD development.

## REFERENCES

1 Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* 1991; **82**: 239–259.
2 Nelson PT, Alafuzoff I, Bigio EH, Bouras C, Braak H, Cairns NJ et al. Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *J Neuropathol Exp Neurol* 2012; **71**: 362–381.
3 Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR, Landfield PW. Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc Natl Acad Sci USA* 2004; **101**: 2173–2178.
4 Wilmot B, McWeeney SK, Nixon RR, Montine TJ, Laut J, Harrington CA et al. Translational gene mapping of cognitive decline. *Neurobiol Aging* 2008; **29**: 524–541.
5 Haroutunian V, Katsel P, Schmeidler J. Transcriptional vulnerability of brain regions in Alzheimer's disease and dementia. *Neurobiol Aging* 2009; **30**: 561–573.
6 Bossers K, Wirz KT, Meerhoff GF, Essing AH, van Dongen JW, Houba P et al. Concerted changes in transcripts in the prefrontal cortex precede neuropathology in Alzheimer's disease. *Brain* 2010; **133**: 3699–3723.
7 Dunckley T, Beach TG, Ramsey KE, Grover A, Mastroeni D, Walker DG et al. Gene expression correlates of neurofibrillary tangles in Alzheimer's disease. *Neurobiol Aging* 2006; **27**: 1359–1371.
8 Murayama S, Saito Y. Neuropathological diagnostic criteria for Alzheimer's disease. *Neuropathology* 2004; **24**: 254–260.
9 Murayama S, Saito Y. The present and future of brain bank in Japan. *Brain Nerve* 2010; **62**: 1013–1018 (in Japanese).
10 Kuwano R, Miyashita A, Arai H, Asada T, Imagawa M, Shoji M et al. Dynamin-binding protein gene on chromosome 10q is associated with late-onset Alzheimer's disease. *Hum Mol Genet* 2006; **15**: 2170–2182.
11 Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 2006; **7**: 3.
12 Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; **4**: 249–264.
13 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Statist Soc Ser B* 1995; **57**: 289–300.

14 Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**: 44–57.

15 Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 1999; **27**: 29–34.

16 Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C et al. The BioGRID interaction database: 2013 update. *Nucleic Acids Res* 2013; **41**: D816–D823.

17 Guillot-Sestier MV, Town T. Innate immunity in Alzheimer's disease: a complex affair. *CNS Neurol Disord Drug Targets* 2013; **12**: 593–607.

18 Oláh J, Vincze O, Virók D, Simon D, Bozsó Z, Tőkési N et al. Interactions of pathological hallmark proteins: tubulin polymerization promoting protein/p25, beta-amyloid, and alpha-synuclein. *J Biol Chem* 2011; **286**: 34088–34100.

19 Bekris LM, Millard S, Lutz F, Li G, Galasko DR, Farlow MR et al. Tau phosphorylation pathway genes and cerebrospinal fluid tau levels in Alzheimer's disease. *Am J Med Genet B Neuropsychiatr Genet* 2012; **159B**: 874–883.

20 Mondragón-Rodríguez S, Perry G, Zhu X, Moreira PI, Acevedo-Aquino MC, Williams S. Phosphorylation of tau protein as the link between oxidative stress, mitochondrial dysfunction, and connectivity failure: implications for Alzheimer's disease. *Oxid Med Cell Longev* 2013; **2013**: 940603.

21 Folsom TD, Fatemi SH. The involvement of Reelin in neurodevelopmental disorders. *Neuropharmacology* 2013; **68**: 122–135.

22 Krstic D, Pfister S, Notter T, Knuesel I. Decisive role of Reelin signaling during early stages of Alzheimer's disease. *Neuroscience* 2013; **246**: 108–116.

23 Minghetti L. Cyclooxygenase-2 (COX-2) in inflammatory and degenerative brain diseases. *J Neuropathol Exp Neurol* 2004; **63**: 901–910.

24 Aïd S, Bosetti F. Targeting cyclooxygenases-1 and -2 in neuroinflammation: therapeutic implications. *Biochimie* 2011; **93**: 46–51.

25 Kamboh MI, Barmada MM, Demirci FY, Minster RL, Carrasquillo MM, Pankratz VS et al. Genome-wide association analysis of age-at-onset in Alzheimer's disease. *Mol Psychiatry* 2012; **17**: 1340–1346.

26 Wang H, Dey D, Carrera I, Minond D, Bianchi E, Xu S et al. COPS5 (Jab1) Protein increases β site processing of amyloid precursor protein and amyloid β peptide generation by stabilizing RanBP9 protein levels. *J Biol Chem* 2013; **288**: 26668–26677.

27 Zhang Z, Kochan GT, Ng SS, Kavanagh KL, Oppermann U, Schofield CJ et al. Crystal structure of PHYHD1A, a 2OG oxygenase related to phytanoyl-CoA hydroxylase. *Biochem Biophys Res Commun* 2011; **408**: 553–558.

28 Loenarz C, Schofield CJ. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat Chem Biol* 2008; **4**: 152–156.

29 Otaegui-Arrazola A, Amiano P, Elbusto A, Urdaneta E, Martínez-Lage P. Diet, cognition, and Alzheimer's disease: food for thought. *Eur J Nutr* 2014; **53**: 1–23.

30 Rodriguez OC, Cheney RE. Human myosin-Vc is a novel class V myosin expressed in epithelial cells. *J Cell Sci* 2002; **115**: 991–1004.

31 Jacobs DT, Weigert R, Grode KD, Donaldson JG, Cheney RE. Myosin Vc is a molecular motor that functions in secretory granule trafficking. *Mol Biol Cell* 2009; **20**: 4471–4488.

32 Grünblatt E, Bartl J, Riederer P. The link between iron, metabolic syndrome, and Alzheimer's disease. *J Neural Transm* 2011; **118**: 371–379.

33 Roland JT, Lapierre LA, Goldenring JR. Alternative splicing in class V myosins determines association with Rab10. *J Biol Chem* 2009; **284**: 1213–1223.

34 Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* 2007; **39**: 168–177.

35 Marx J. Molecular biology. Trafficking protein suspected in Alzheimer's disease. *Science* 2007; **315**: 314.

36 Carpenter S, Carlson T, Dellacasagrande J, Garcia A, Gibbons S, Hertzog P et al. TRIL, a functional component of the TLR4 signaling complex, highly expressed in brain. *J Immunol* 2009; **183**: 3989–3995.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)