Original Investigation

# A Multiancestral Genome-Wide Exome Array Study of Alzheimer Disease, Frontotemporal Dementia, and Progressive Supranuclear Palsy

Jason A. Chen, BSE; Qing Wang, PhD; Jeremy Davis-Turak, BA; Yun Li, PhD; Anna M. Karydas, BA; Sandy C. Hsu, MS; Renee L. Sears, BA; Doxa Chatzopoulou, MS; Alden Y. Huang, BS; Kevin J. Wojta, BS; Eric Klein, MSc; Jason Lee, BS; Duane L. Beekly, BS; Adam Boxer, MD, PhD; Kelley M. Faber, MS; Claudia M. Haase, PhD; Josh Miller, PhD; Wayne W. Poon, PhD; Ami Rosen, MS; Howard Rosen, MD; Anna Sapozhnikova, BA; Jill Shapira, RN, PhD; Arousiak Varpetian, MD; Tatiana M. Foroud, PhD; Robert W. Levenson, PhD; Allan I. Levey, MD, PhD; Walter A. Kukull, PhD; Mario F. Mendez, MD, PhD; John Ringman, MD; Helena Chui, MD; Carl Cotman, PhD; Charles DeCarli, MD; Bruce L. Miller, MD; Daniel H. Geschwind, MD, PhD; Giovanni Coppola, MD

**IMPORTANCE** Previous studies have indicated a heritable component of the etiology of neurodegenerative diseases such as Alzheimer disease (AD), frontotemporal dementia (FTD), and progressive supranuclear palsy (PSP). However, few have examined the contribution of low-frequency coding variants on a genome-wide level.

**OBJECTIVE** To identify low-frequency coding variants that affect susceptibility to AD, FTD, and PSP.

**DESIGN, SETTING, AND PARTICIPANTS** We used the Illumina HumanExome BeadChip array to genotype a large number of variants (most of which are low-frequency coding variants) in a cohort of patients with neurodegenerative disease (224 with AD, 168 with FTD, and 48 with PSP) and in 224 control individuals without dementia enrolled between 2005-2012 from multiple centers participating in the Genetic Investigation in Frontotemporal Dementia and Alzheimer's Disease (GIFT) Study. An additional multiancestral replication cohort of 240 patients with AD and 240 controls without dementia was used to validate suggestive findings. Variant-level association testing and gene-based testing were performed.

**MAIN OUTCOMES AND MEASURES** Statistical association of genetic variants with clinical diagnosis of AD, FTD, and PSP.

**RESULTS** Genetic variants typed by the exome array explained 44%, 53%, and 57% of the total phenotypic variance of AD, FTD, and PSP, respectively. An association with the known AD gene *ABCA7* was replicated in several ancestries (discovery $P = .0049$, European $P = .041$, African American $P = .043$, and Asian $P = .027$), suggesting that exonic variants within this gene modify AD susceptibility. In addition, 2 suggestive candidate genes, *DYSF* ($P = 5.53 \times 10^{-5}$) and *PAXIP1* ($P = 2.26 \times 10^{-4}$), were highlighted in patients with AD and differentially expressed in AD brain. Corroborating evidence from other exome array studies and gene expression data points toward potential involvement of these genes in the pathogenesis of AD.

**CONCLUSIONS AND RELEVANCE** Low-frequency coding variants with intermediate effect size may account for a significant fraction of the genetic susceptibility to AD and FTD. Furthermore, we found evidence that coding variants in the known susceptibility gene *ABCA7*, as well as candidate genes *DYSF* and *PAXIP1*, confer risk for AD.

*JAMA Neurol*. doi:10.1001/jamaneurol.2014.4040
Published online February 23, 2015.

Editorial

Supplemental content at jamaneurology.com

**Author Affiliations:** Author affiliations are listed at the end of this article.

**Corresponding Author:** Giovanni Coppola, MD, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, 3506C Gonda Neuroscience and Genetics Research Center, 695 Charles E. Young Dr S, Los Angeles, CA 90095 (gcoppola@ucla.edu).

Genetics studies have revealed a genetic contribution to susceptibility for common or sporadic forms of neurodegenerative disease such as Alzheimer disease (AD), frontotemporal dementia (FTD), and progressive supranuclear palsy (PSP, a syndrome characterized by oculomotor and gait abnormalities). In AD, early genetic mapping approaches have identified rare variants in genes such as *APP*, *PSEN1*, and *PSEN2* that cause familial, early-onset forms.[1] *APOE* was also pinpointed as a late-onset AD susceptibility gene.[2] Genome-wide association studies[3-5] (GWAS) targeted toward common variants in primarily European populations have identified many variants associated with AD, most clearly near *APOE* but also consistently near *ABCA7*, *BIN1*, *CLU*, *CR1*, *PICALM*, *SORL1*, and other genes. Next-generation sequencing approaches have also found rare variants with strong effect in the *MAPT* and *TREM2* genes.[6,7]

In FTD, the most frequently observed mutations in familial cases occur in *C9ORF72*, *GRN*, *MAPT*, *TARDBP*, and other genes.[8] In sporadic cases, a haplotype variant on the long arm of chromosome 17 has been repeatedly associated with PSP.[9-11] In addition, GWAS have been performed for sporadic cases of FTD, identifying associated single-nucleotide polymorphisms (SNPs) near *TMEM106B*[12] and *BTNL2*/*HLA-DRA*/*HLA-DRB5* and *RAB38*/*CTSC*,[13] as well as for PSP, identifying associated SNPs near *MAPT*, *EIF2AK3*, *STX6*, and *MOBP*.[11]

Despite progress in understanding the genetics of neurodegenerative diseases, known genetic risk factors cannot explain a large portion of the heritability of these diseases. For example, in AD, all common variants (including known and unknown risk variants) have been predicted to account for less than 25% of disease variance,[14] and known high-penetrance rare variants account for few cases, collectively totaling only a fraction of the estimated 58% to 79% heritability of AD.[15] Some of this missing heritability may be due to a blind spot in conventional genetic studies to date. A moderately rare variant with moderate effect size would be too uncommon to be tagged by a standard genotyping array and have too small of an effect to be detected by linkage or genome sequencing in practical sample sizes. The exome array bridges this gap by genotyping at low cost more than 200 000 coding variants identified through sequencing studies (**Figure 1**). This approach has been applied to phenotypes such as insulin homeostasis,[16] bronchopulmonary dysplasia,[17] and heart disease.[18,19] For AD, Chung et al[20] recently reported an exome array study in Korean participants that found an association with *APOE*, *APOC1*, and *TOMM40* variants (near the *APOE* locus) but did not identify novel genetic variants. Herein, we report findings from the application of the exome array to the multiancestral Genetic Investigation in Frontotemporal Dementia and Alzheimer's Disease (GIFT) Study cohort to determine the contribution of low-frequency coding variants to susceptibility to sporadic AD, PSP, and FTD.

## Methods

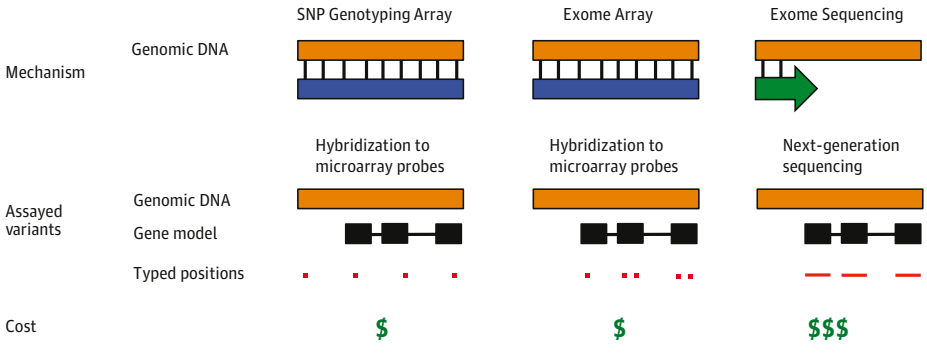
### Study Cohort

Patients and healthy control individuals were enrolled between 2005-2012 at the Memory and Aging Center, University of California, San Francisco, as part of the GIFT Study, an investigation of the genetics of neurodegenerative disease .[21,22] Written consent was obtained at the participating institutions. The study was approved by the Institutional Review Board of the University of California, Los Angeles. An additional 32 DNA samples from patients with PSP were extracted from postmortem brain tissue from the New York Brain Bank at Columbia University (New York, New York). A subset of these individuals were initially selected for genotyping using the Illumina HumanExome BeadChip array (**Table 1**). Patients diagnosed as having FTD with motor neuron disease (FTD/MND) were excluded from further analysis owing to the small sample size and potential genetic heterogeneity.

### Replication Cohort

As part of the GIFT Study, individuals were also enrolled from other sites, including Emory University, University of Southern California, and University of California at Berkeley, Davis, Irvine, and Los Angeles. Following initial data analysis, 480 individuals from this additional group of patients, including 240 patients with AD and 240 controls without dementia, were genotyped (**Table 2**). These individuals were analyzed as above but owing to genetic heterogeneity were divided into 4 general groups (European, African American, Latino, and Asian) based on self-reported ancestry. To ensure proper classification and minimize the inclusion of misplated samples, genetic ancestry was also estimated by multidimensional

Figure 1. Comparison of the Exome Array and Related Genotyping and Sequencing Technologies




The exome array serves as a bridge between conventional single-nucleotide polymorphism (SNP) genotyping array and exome sequencing. The exome array assays primarily variants within exonic regions of the DNA, similar to exome sequencing; however, the location of the variants must be known a priori. The cost of the exome array is typically similar to that of other genotyping arrays and is much less expensive than that of exome sequencing.

Table 1. Demographic Information for the Discovery Cohort

| Characteristic | AD (n = 224) | Control (n = 224) | FTD (n = 168) | FTD/MND (n = 8) | PSP (n = 48) |
|---|---|---|---|---|---|
| Age, median (range), y | 71 (42 to ≥89) | 71 (35 to ≥89) | 67 (35 to ≥89) | 63 (35 to 80) | 76 (55 to ≥89) |
| Sex, No. (%) | | | | | |
| Male | 12 (56.7) | 94 (42.0) | 95 (56.5) | 8 (100) | 19 (39.6) |
| Female | 97 (43.3) | 130 (58.0) | 73 (43.5) | 0 | 29 (60.4) |
| Ancestry, No. (%) | | | | | |
| European | 195 (87.1) | 183 (81.7) | 144 (85.7) | 8 (100) | 12 (25.0) |
| African American | 2 (0.9) | 0 | 0 | 0 | 0 |
| Latino | 0 | 4 (1.8) | 1 (0.6) | 0 | 0 |
| Asian | 20 (8.9) | 27 (12.1) | 9 (5.4) | 0 | 1 (2.1) |
| Other | 3 (1.3) | 4 (1.8) | 7 (4.2) | 0 | 2 (4.2) |
| Unknown | 4 (1.8) | 6 (2.7) | 7 (4.2) | 0 | 33 (68.8) |
| *APOE* genotype, No. (%) | | | | | |
| *E2/E2* | 1 (0.4) | 1 (0.4) | 0 | 0 | 1 (2.1) |
| *E2/E3* | 7 (3.1) | 19 (8.5) | 16 (9.5) | 1 (12.5) | 3 (6.3) |
| *E2/E4* | 4 (1.8) | 1 (0.4) | 1 (0.6) | 0 | 2 (4.2) |
| *E3/E3* | 99 (44.2) | 157 (70.1) | 107 (63.7) | 5 (62.5) | 36 (75.0) |
| *E3/E4* | 92 (41.1) | 40 (17.9) | 40 (23.8) | 1 (12.5) | 6 (12.5) |
| *E4/E4* | 21 (9.4) | 6 (2.7) | 4 (2.4) | 1 (12.5) | 0 |
| Chromosome 17q21.31 haplotype, No. (%) | | | | | |
| *H1/H1* | 91 (40.6) | 132 (58.9) | 107 (63.7) | 4 (50.0) | 43 (89.6) |
| *H1/H2* | 48 (21.4) | 52 (23.2) | 33 (19.6) | 3 (37.5) | 5 (10.4) |
| *H2/H2* | 4 (1.8) | 10 (4.5) | 7 (4.2) | 0 | 0 |
| Untyped | 81 (36.2) | 30 (13.4) | 21 (12.5) | 1 (12.5) | 0 |

Abbreviations: AD, Alzheimer disease; FTD, frontotemporal dementia; FTD/MND, FTD with motor neuron disease; PSP, progressive supranuclear palsy.

Table 2. Demographic Information for the Replication Cohort

| Characteristic | No. (%) | | | |
|---|---|---|---|---|
| | European (n = 135) | African American (n = 271) | Latino (n = 50) | Asian (n = 24) |
| Diagnosis | | | | |
| AD | 68 (50.4) | 138 (50.9) | 21 (42.0) | 13 (54.2) |
| Control | 67 (49.6) | 133 (49.1) | 29 (58.0) | 11 (45.8) |
| Sex | | | | |
| Male | 68 (50.4) | 73 (26.9) | 19 (38.0) | 8 (33.3) |
| Female | 57 (42.2) | 198 (73.1) | 31 (62.0) | 16 (66.7) |
| Unknown | 10 (7.4) | 0 | 0 | 0 |
| Contributing center | | | | |
| Emory University | 21 (15.6) | 223 (82.3) | 0 | 0 |
| University of California, Berkeley | 33 (24.4) | 14 (5.2) | 8 (16.0) | 8 (33.3) |
| University of California, Davis | 3 (2.2) | 32 (11.8) | 23 (46.0) | 5 (20.8) |
| University of California, Irvine | 55 (40.7) | 2 (0.7) | 5 (10.0) | 1 (4.2) |
| University of California, Los Angeles | 2 (1.5) | 0 | 0 | 0 |
| University of California, San Francisco | 20 (14.8) | 0 | 0 | 6 (25.0) |
| University of Southern California | 1 (0.7) | 0 | 14 (28.0) | 4 (16.7) |
| *APOE* genotype | | | | |
| *E2/E2* | 1 (0.7) | 2 (0.7) | 0 | 0 |
| *E2/E3* | 4 (3.0) | 16 (5.9) | 2 (4.0) | 1 (4.2) |
| *E2/E4* | 5 (3.7) | 9 (3.3) | 3 (6.0) | 0 |
| *E3/E3* | 41 (30.4) | 87 (32.1) | 34 (68.0) | 8 (33.3) |
| *E3/E4* | 21 (15.6) | 86 (31.7) | 9 (18.0) | 3 (12.5) |
| *E4/E4* | 9 (6.7) | 12 (4.4) | 1 (2.0) | 2 (8.3) |
| Untyped | 54 (40.0) | 59 (21.8) | 1 (2.0) | 10 (41.7) |

Abbreviation: AD, Alzheimer disease.

Table 3. GCTA Explained Variance Analysis

| Variable | Variance Explained (SE) | | |
|---|---|---|---|
| | AD | FTD | PSP |
| All exome array variants[a] | 0.44 (0.39) | 0.53 (0.36) | 0.57 (0.44) |
| Exonic fraction | 0.50 (0.36) | 0.45 (0.35) | 0.26 (0.56) |
| Low-frequency exonic fraction[b] | 0.41 (0.39) | 0.42 (0.37) | 0.03 (0.58) |

Abbreviations: AD, Alzheimer disease; FTD, frontotemporal dementia; GCTA, Genome-Wide Complex Trait Analysis (http://www.complextraitgenomics.com/software/gcta/); PSP, progressive supranuclear palsy.

[a] Includes genome-wide association studies hits, HLA tag single-nucleotide polymorphisms, custom content, ancestry-informative single-nucleotide polymorphisms, and others.

[b] Less than 5% minor allele frequency between all disease cohorts and control subjects.

scaling using the PLINK whole-genome association analysis tool set (http://pngu.mgh.harvard.edu/purcell/plink/) using the entire set of genotyped variants by the exome array. Following this procedure, 44 samples were suspected of misclassification and were removed from further analysis.

### Exome Array Genotyping

Exonic and nonexonic variants were genotyped using the Illumina Infinium HumanExome BeadChip kit. While mostly consisting of coding variants from prior sequencing studies, the exome arrays also included markers for previously described GWAS hits, ancestry-informative markers, randomly selected synonymous variants, HLA tag SNPs, and others,[16] in total comprising 250 272 genotyped markers per sample. Quality control procedures were enacted to remove suspect variants and minimize the effect of population structure on the data analysis. The eMethods, eFigure 1, and eFigure 2 in the Supplement provide further details on genotyping and data preprocessing procedures.

### Statistical Analysis

The total phenotypic (disease) variance explained by the genotyped variants was determined using a restricted maximum likelihood model implemented in Genome-Wide Complex Trait Analysis (GCTA; http://www.complextraitgenomics.com/software/gcta/). Variant-level association with AD, FTD, and PSP was tested using a logistic regression model that corrected for population structure. The association on the gene level was tested using the sequence kernel association test (SKAT),[23] a nonburden test that is sensitive in the presence of neutral genetic variants. Genes that showed suggestive associations with AD were also tested in previously described brain messenger RNA (mRNA) expression data sets.[24,25] The eMethods in the Supplement provides a more detailed description of the statistical methods used.

Summary statistics and individual-level data are available from the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS; https://www.niagads.org/, accession number NG00040).

## Results

### Patient Characteristics

The initial discovery sample included 224 patients with AD, 168 patients with FTD, 8 patients with FTD/MND, 48 patients with PSP, and 224 healthy controls. Demographic characteristics are summarized in Table 1. The ancestral makeup of this sample was predominantly European (80.7% overall). Consistent with their known roles in the respective diseases, individuals classified as having AD showed high prevalence of the *APOE ε4* allele (41.1% *ε3/ε4* and 9.4% *ε4/ε4*), and individuals classified as having PSP showed high prevalence of the *H1* haplotype (89.6% *H1/H1* and 10.4% *H1/H2*). The replication cohort consisted of a more ancestrally heterogeneous set of patients and controls (Table 2).

### Low-Frequency Exonic Variants Explain a Fraction of the Phenotypic Variation in AD and FTD

For each of the 3 diseases (AD, FTD, and PSP), the GCTA software was applied to the data set to estimate the variance explained by the following 3 different classes of variants: all variants, including nonexonic variants; exonic variants only; and low-frequency exonic variants, with minor allele frequency <5%. In each case, a substantial portion of the observed phenotypic variance could be explained by all the typed variants (**Table 3**). However, owing to the small sample sizes on which each of these estimates is based, the standard error of each measurement is high.

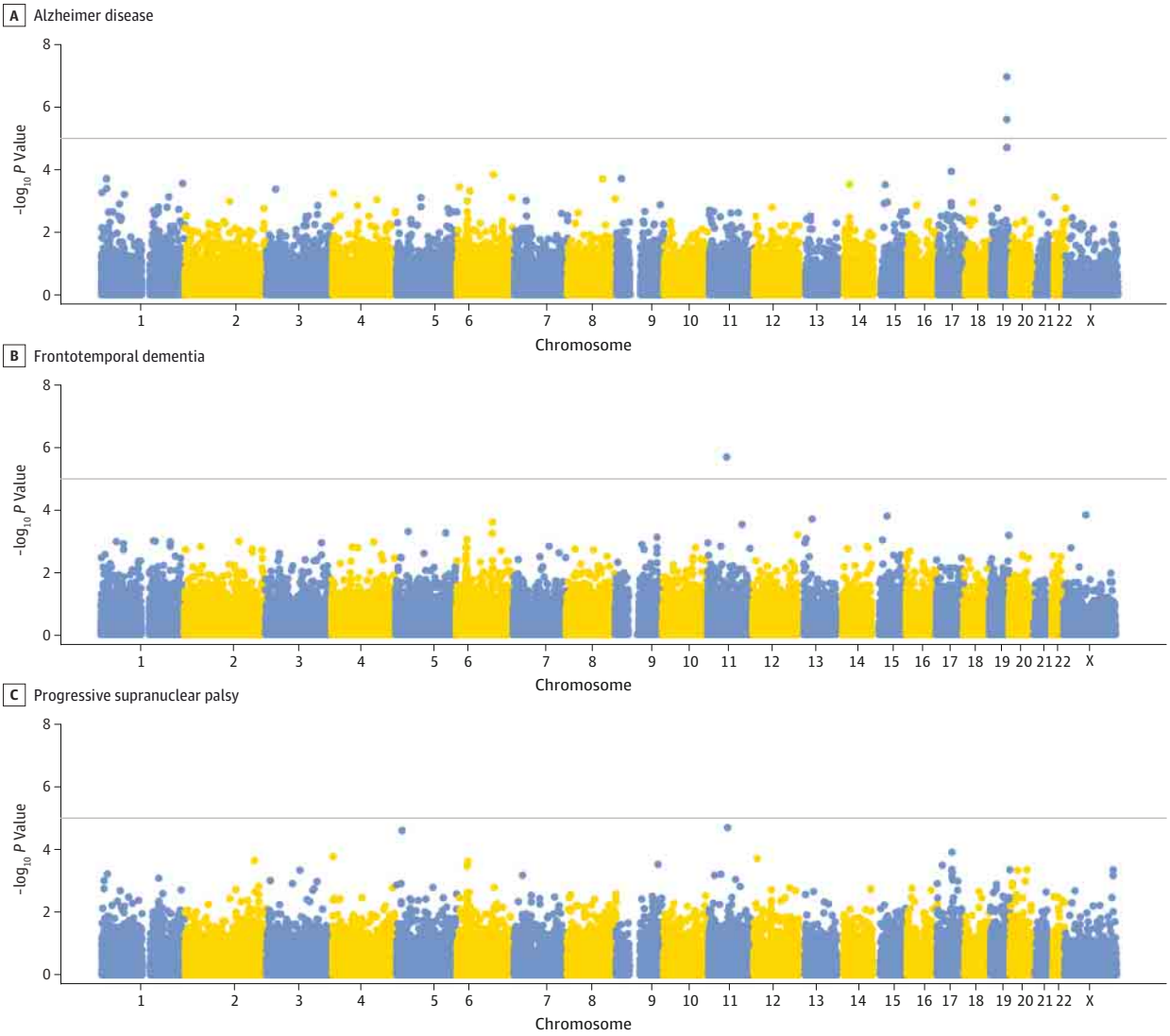
### Variant-Level Association Testing Identifies Significant Associations With Known and Novel Loci

A logistic regression procedure was performed on our discovery cohort to test for an association with AD, FTD, or PSP. Our method largely controlled for genomic inflation due to population stratification in each of the 3 disease categories (eFigure 3 in the Supplement). Two variants were suggestively associated with AD, rs769449 ($P = 1.14 \times 10^{-7}$; minor allele odds ratio [OR], 3.0) and rs4420638 ($P = 2.58 \times 10^{-6}$; minor allele OR, 2.3). Both variants are within the *APOE/TOMM40/APOC1* region on chromosome 19 identified in previous genetic studies.[2-5] One variant was associated with FTD, exm2250002 ($P = 2.08 \times 10^{-6}$; minor allele OR, 0.8), corresponding to a synonymous exonic variant in the olfactory receptor genes *OR9G1* and *OR9G9*. No variants reached the suggestive *P* value threshold ($1 \times 10^{-5}$) in the PSP cohort. Manhattan plots depicting associations in AD, FTD, and PSP are shown in **Figure 2**.

### Exome Array Genotyping Replicates Some Previous Associations Found in AD and PSP

Thirty-nine polymorphisms previously associated with AD and 9 polymorphisms associated with PSP (National Human

Figure 2. Manhattan Plot of Associations in Alzheimer Disease, Frontotemporal Dementia, and Progressive Supranuclear Palsy




The association $-\log_{10} P$ values calculated by logistic regression are presentd for for Alzheimer disease, frontotemporal dementia, and progressive supranuclear palsy. The horizontal line indicates the suggestive $P$ value threshold of $P = 1 \times 10^{-5}$. X refers to chromosome X.

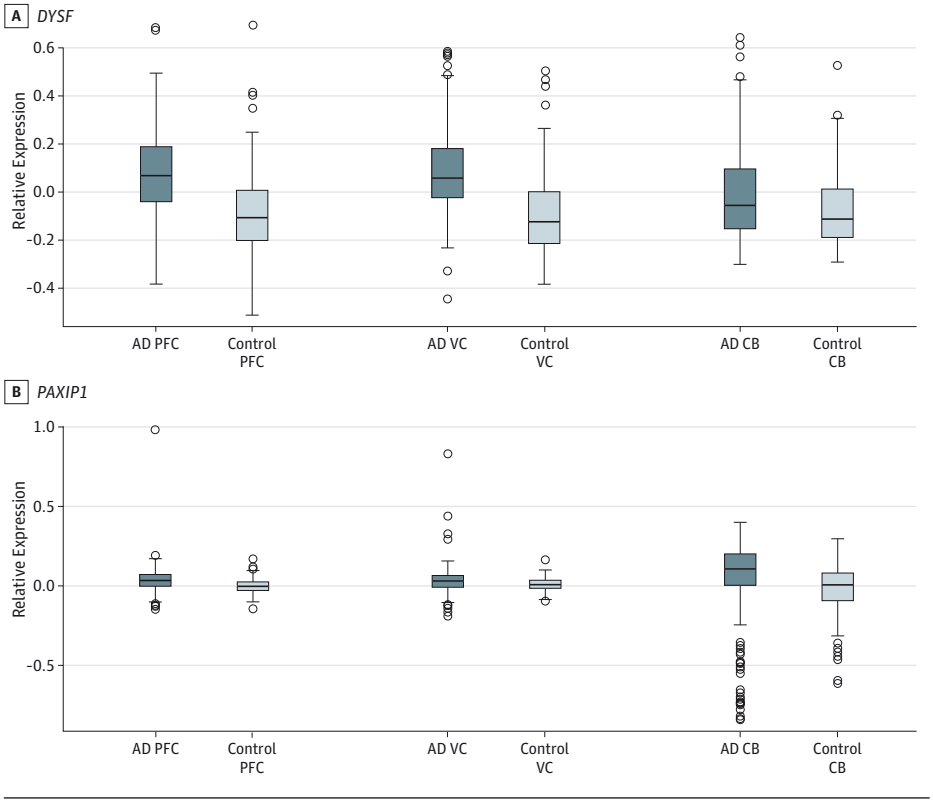
Genome Research Institute Genome-Wide Association Studies Catalog; http://www.genome.gov/gwastudies/) were typed by the exome array. Reported susceptibility loci for FTD were not typed on this platform. We tested the association between each of these variants and their respective disease in our cohort, as calculated by the logistic procedure described previously. For AD, the Bonferroni correction for 39 tests at a familywise error rate of .05 yielded a $P$ value threshold at .0013. Two associations near *APOE*, rs2075650 ($P = 2.05 \times 10^{-5}$) and rs4420638 ($P = 2.58 \times 10^{-6}$), surpassed this predefined $P$ value threshold (eTable 1 in the Supplement). While the other tested GWAS variants were not significantly associated with AD, the overall direction of the association was highly consistent with previously reported results,[3-5] and 23 of 32 SNPs for which the risk allele was unambiguous showed the same direction of effect as previously reported ($P = .010$, binomial test).

For PSP, the Bonferroni correction for 9 tests at a familywise error rate of .05 yielded a $P$ value threshold at .0056. A single variant exceeded this threshold, rs8070723 ($P = .00043$) on chromosome 17 near *MAPT* (eTable 2 in the Supplement). Similar to the AD cohort, the direction of the association was highly consistent with previously reported results,[11] with 8 of 9 SNPs showing the same direction of effect ($P = .019$, binomial test).

## Gene-Level Testing Suggests Several AD Candidate Genes

Gene-level hypothesis testing was performed using SKAT-derived $P$ values for 17 141 genes (that contained at least 1 typed variant after quality control). Using a permutation procedure, a false discovery rate of 50% was expected to be controlled at a SKAT-derived $P$ value of $4.54 \times 10^{-4}$ for AD, $5.06 \times 10^{-4}$ for FTD, and $9.65 \times 10^{-5}$ for PSP. For AD, the following 6 genes exceeded

Figure 3. Differential Expression of *DYSF* and *PAXIP1* in Alzheimer Disease (AD) Brain




Shown is the expression of *DYSF* (A) and *PAXIP1* (B) in a public microarray data set of brain messenger RNA, grouped by brain region, in patients with AD (dark gray) vs healthy control subjects without dementia (light gray). The vertical axis represents the normalized expression residual, corrected for technical covariates. CB indicates cerebellum; PFC, prefrontal cortex; and VC, visual cortex.

this threshold: *DYSF*, *PAXIP1*, *TOP1MT*, *C3ORF1*, *SETDB1*, and *CRISPLD1* ($P = 5.53 \times 10^{-5}$, $P = 2.26 \times 10^{-4}$, $P = 2.29 \times 10^{-4}$, $P = 3.93 \times 10^{-4}$, $P = 4.13 \times 10^{-4}$, and $P = 4.54 \times 10^{-4}$, respectively). For FTD, the following 8 genes exceeded the threshold: *RAB21*, *AKR1B10*, *C9ORF6*, *CD5L*, *WDR38*, *OPHN1*, *ADORA3*, and *IKBKAP* ($P = 4.65 \times 10^{-5}$, $P = 4.83 \times 10^{-5}$, $P = 2.55 \times 10^{-4}$, $P = 3.65 \times 10^{-4}$, $P = 3.85 \times 10^{-4}$, $P = 4.78 \times 10^{-4}$, $P = 4.79 \times 10^{-4}$, and $P = 5.06 \times 10^{-4}$, respectively). For PSP, 2 genes exceeded the threshold, *OR1Q1* and *VWA3A* ($P = 3.00 \times 10^{-5}$ and $P = 9.65 \times 10^{-5}$, respectively).

We attempted to replicate the findings for AD in an additional multiancestral cohort of 240 cases and 240 controls. No further samples from patients with FTD or PSP were available, so those results could not be tested. Using the Bonferroni correction, a *P* value threshold of .0021 (considering 6 genes times 4 ancestry categories, for a total of 24 tests) was determined to control for a familywise error rate of .05. None of the suggestive genes identified for AD were significant under this threshold in any ancestral category in the replication cohort (eTable 3 in the Supplement). However, several genes trended toward significance in some cases, including *DYSF* in Europeans ($P = .076$), *PAXIP1* in Latinos and Asians ($P = .016$ and $P = .037$, respectively), and *TOP1MT* in African Americans ($P = .0059$). Because of previous reports of the involvement of *DYSF* and *PAXIP1* in the AD literature (see the Discussion section below),[26,27] these genes were considered interesting candidate genes for AD susceptibility. Overall, we analyzed 38 variants in *DYSF* (including 3 synonymous and 35 missense) and 5 variants in *PAXIP1* (including 1 synonymous and 4 missense) typed by the exome array, demonstrating variation in our cohort, and passing quality control criteria.

We further identified 71 genes previously implicated in genetic studies of AD as categorized in the Human Gene Mutation Database[28] (version 2014.1) and extracted the association statistics in the initial discovery set and the 4 replication cohorts. Only *ABCA7* (OMIM 605414) (SKAT discovery $P = .0049$) reached nominal significance. Notably, the SKAT *P* value was also nominally significant in the European ($P = .041$), African American ($P = .043$), and Asian ($P = .027$) replication cohorts but not the Latino ($P = .61$) cohort.

## *DYSF* and *PAXIP1* Transcripts Are Differentially Expressed in AD Brain

To further solidify whether *DYSF* (OMIM 603009) and *PAXIP1* (OMIM 608254) are involved in the pathogenesis of AD, we examined their relative expression levels in patients with AD and controls without dementia in a published microarray data set.[24] The expression of *DYSF* and *PAXIP1* was significantly different between cases and controls in each of the examined brain regions (**Figure 3**). In the prefrontal cortex, visual cortex, and cerebellum, the expression of *DYSF* was increased in patients with AD ($P < 2.2 \times 10^{-16}$, $P = 2.33 \times 10^{-15}$, and $P = .00080$, respectively). These findings were corroborated by independent data,[25] which also showed increased expression of *DYSF* in the cerebral cortex of patients with AD ($P = .00023$). Similarly, the expression of *PAXIP1* in the prefrontal cortex, visual

cortex, and cerebellum was increased in patients with AD ($P = 3.6 \times 10^{-14}$, $P = .0034$, and $P = .00095$, respectively).

## Discussion

We evaluated the contribution of exonic variants to neurodegenerative disease susceptibility in a multiancestral cohort totaling 464 patients with AD, 168 patients with FTD, 48 patients with PSP, and 464 controls without dementia. We found that low-frequency (<5%) coding variants explain a sizable proportion of the phenotypic variance in AD and FTD, although the confidence limits for this estimate are large owing to our sample size. Well-known associations with the *APOE* locus for AD and 17q21.31 haplotype for PSP were replicated, and a novel susceptibility locus was identified at exm2250002 for FTD. Whether this variant is a true genetic signal is questionable given that it was also the most significant signal in the PSP cohort ($P = 2.03 \times 10^{-5}$) and corresponds to a synonymous variant within *OR9G1*/*OR9G9*, members of the polymorphic olfactory receptor family. Gene-level testing identified suggestive signals from *DYSF* and *PAXIP1* in AD, and a trend toward significance was observed in a replication cohort in several of the tested ancestral categories. A possible contribution to disease risk from exonic variants in the AD susceptibility gene *ABCA7* was also detected in multiple ancestral categories. However, we caution that these results are merely suggestive and await validation in well-powered cohorts and model systems.

The focus of the exome array on coding variation, much of which has low frequency in the general population, means that large sample sizes are needed to observe statistically significant effects, unless the effect sizes are large, as is the case with the association of the *APOE* ε4 allele with AD. We estimated that a variant at 5% minor allele frequency must have a greater than 4-fold OR to achieve 80% power to identify in our AD discovery cohort. Therefore, our initial cohort of 672 patients and controls and our follow-up cohort of 480 patients and controls are underpowered to detect associations with rare variants of modest or intermediate effect sizes. Taken together with heritability estimates, our analyses indicated that rare variants of low or modest effect have a role in AD, FTD, and PSP, late-onset diseases for which deleterious alleles are presumably under weak selective pressure.

Furthermore, while the GIFT Study cohort enabled testing of an association in multiple ancestral groups simultaneously, our results were limited by the small sample sizes. Therefore, our findings do not exclude the possibility that exonic variants with lower frequency or effect size are present in the general population. In fact, the strong association with *ABCA7* (a GWAS-implicated AD susceptibility gene) by SKAT in several ancestral populations strongly suggests that coding variants of modest effect size within this gene are associated with AD risk. Previous GWAS have reported associations with intronic polymorphisms such as rs4147929,[5] rs115550680,[29] and rs3764650,[4] as well as the missense polymorphism rs3752246.[3] It is possible that these variants may tag haplotypes containing causal, exonic variants. Therefore, it is reasonable to attempt to identify novel candidate genes containing multiple, low-frequency coding variants that may contribute to AD.

While not strictly genome-wide significant, genewise testing results reinforce prior findings that have implicated both *DYSF* and *PAXIP1* in the pathogenesis of AD. *DYSF* encodes the protein dysferlin, and mutations in this gene are known to cause autosomal recessive muscle diseases such as Miyoshi myopathy[30] and limb-girdle muscular dystrophy type 2B,[31] known as dysferlinopathies. In skeletal muscle, dysferlin is thought to have a role in calcium-dependent sarcolemma repair.[32,33] Although its function in the central nervous system has not been extensively elaborated, dysferlin has been shown to accumulate in endothelial cells near multiple sclerosis lesions[34] and within Aβ plaques of patients with AD.[26] The colocalization of dysferlin and Aβ42 aggregates was also demonstrated in sporadic inclusion body myositis, suggesting that Aβ may sequester dysferlin and interfere with its normal repair functions in skeletal muscle.[35]

The second highlighted gene, *PAXIP1*, encodes for a nuclear protein with 6 BRCT domains, hinting at its function in DNA repair pathways.[36] PAXIP1 may participate in p53 activation mediated by the ataxia-telangiectasia mutated (ATM) serine/threonine kinase.[36-38] Although variants in *PAXIP1* have not been definitively associated with disease, Rademakers et al[27] identified a significant linkage peak at 7q36 in a large pedigree with multiplex AD. The risk allele of the *D7S798* marker also appeared to increase AD risk by 2.7 times in a Dutch population-based cohort.[27] Sequencing of the coding exons of 29 candidate genes revealed only a single rare variant, a synonymous Ala626 change in *PAXIP1*.

To our knowledge, the neuropathological findings by Galvin et al[26] and the linkage study by Rademakers et al[27] are the only publications to date that implicate *DYSF* and *PAXIP1* in the pathogenesis of AD. Our analysis of published microarray studies indicated increases in *DYSF* and *PAXIP1* mRNA expression in brain regions of patients with AD. However, these results do not provide direct evidence of the roles of these genes in AD. In contrast, the exome array results add additional support for the causal pathogenicity of *DYSF* and *PAXIP1*. Although we could not ascertain whether any of the assayed variants directly affected the expression of *DYSF* and *PAXIP1*, the fact that these genes were both identified by exome array analysis and by differential expression analysis provides convergent evidence for their involvement in AD. Besides partial, nominal replication within our cohort, our findings are further corroborated by a recently published exome array study[20] in AD reporting a strong (but not genome-wide significant) association for *DYSF* ($P = 1.6 \times 10^{-5}$) with AD in a Korean cohort; the association with *PAXIP1* was not reported. The overlap with our suggestive results indicates a high prior probability for the pathogenicity of variants in *DYSF* (and possibly *PAXIP1*), and follow-up studies are warranted.

## Conclusions

The overall genetic architecture of neurodegenerative diseases is complex and is just beginning to be defined. Our work

has strengthened the case for 2 AD candidate genes and provides one of the first glimpses at this genetic variation that heretofore had not been widely studied. We anticipate that the results described herein will provide insight into the genetics of AD, FTD, and PSP and that the data will provide a valuable multiancestral cohort with exome array genotyping data for future studies in each of the 3 diseases. We further expect in the long term that increased understanding of the genetic underpinnings will lead to improvements in diagnosis and management for patients with neurodegenerative diseases.

## ARTICLE INFORMATION

**Accepted for Publication:** November 3, 2014.

**Published Online:** February 23, 2015. doi:10.1001/jamaneurol.2014.4040.

**Author Affiliations:** Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles (Chen, Wang, Davis-Turak, Li, Hsu, Sears, Chatzopoulou, Huang, Wojta, Klein, Lee, Geschwind, Coppola); Memory and Aging Center, University of California, San Francisco (Karydas, Boxer, H. Rosen, B. L. Miller); National Alzheimer's Coordinating Center, University of Washington, Seattle (Beekly, Kukull); Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis (Faber, Foroud); Department of Psychology, School of Education and Social Policy, Northwestern University, Evanston, Illinois (Haase); Department of Nutritional Sciences, Rutgers University, New Brunswick, New Jersey (J. Miller); Institute for Memory Impairments and Neurological Disorders, University of California, Irvine (Poon, Cotman); Department of Neurology, Emory University, Atlanta, Georgia (A. Rosen, Levey); Department of Psychology, University of California, Berkeley (Sapozhnikova, Levenson); Department of Neurology, University of California, Los Angeles (Shapira, Mendez, Ringman, Geschwind, Coppola); Department of Neurology, University of Southern California, Los Angeles (Varpetian, Chui); Mary S. Easton Center for Alzheimer's Disease Research at UCLA, University of California, Los Angeles (Ringman); Department of Neurology, University of California, Davis (DeCarli).

**Author Contributions:** Dr Coppola had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
*Study concept and design:* Chen, Li, Miller, Geschwind, Coppola.
*Acquisition, analysis, or interpretation of data:* All authors.
*Drafting of the manuscript:* Chen, Geschwind, Coppola.
*Critical revision of the manuscript for important intellectual content:* All authors.
*Statistical analysis:* Chen, Li, Coppola.
*Obtained funding:* Levenson, Ringman, DeCarli, Miller, Geschwind, Coppola.
*Administrative, technical, or material support:* Geschwind, Coppola.
*Study supervision:* Geschwind, Coppola.

**Conflict of Interest Disclosures:** None reported.

**Funding/Support:** This work was funded by grants from the National Institutes of Health: F31 NS084556 (Mr Chen), P50 AG023501 (Dr Miller), RC1 AG035610 (Dr Coppola), R01 MH097268 (Dr Coppola), R01 AG26938 (Dr Geschwind), P01 AG019724 (Drs Levenson, B. L. Miller, and Geschwind), and R01 AG041762 (Dr Levenson); from the John Douglas French Alzheimer's Foundation (Dr Coppola); the Tau Consortium (Drs Geschwind and Coppola); and the Jim Easton Consortium for Alzheimer's Drug Discovery and Biomarker Development (Dr Ringman). We acknowledge the support of grant P30 NS062691 from the National Institute of Neurological Disorders and Stroke Informatics Center for Neurogenetics and Neurogenomics; grant P50 AG16570 from the University of California, Los Angeles, Alzheimer's Disease Research Center; grant P50 AG016573 and POI AG000538 from the University of California, Irvine, Alzheimer's Disease Research Center; grant P30 AG010129 from the University of California, Davis, Alzheimer's Disease Center (Dr DeCarli); and grant P50 AG05142 from the University of Southern California Alzheimer's Disease Research Center. Samples from the National Cell Repository for Alzheimer's Disease, which receives government support under cooperative agreement grant U24 AG21886 from the National Institute on Aging, were used in this study. The National Alzheimer's Coordinating Center database is funded by grant U01 AG016979 from the National Institute on Aging.

**Role of the Funder/Sponsor:** The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Additional Contributions:** Iris Vuong, BS, and Joseph DeYoung, BA (University of California, Los Angeles Neurosciences Genomics Core) assisted with array genotyping. Jean Paul Vonsattel, MD (New York Brain Bank at Columbia University) contributed brain specimens from patients with PSP that were used in the study. Bin Zhang, PhD, and Amanda Myers, PhD, assisted with the gene expression data. Margaret Chu, BS, and Maribel Estrada provided administrative support. We thank the contributors who collected samples used in this study. We especially thank the patients and their families, whose help and participation made this work possible and to whom our research is dedicated.

## REFERENCES

**1**. Lendon CL, Ashall F, Goate AM. Exploring the etiology of Alzheimer disease using molecular genetics. *JAMA*. 1997;277(10):825-831.

**2**. Pericak-Vance MA, Bebout JL, Gaskell PC Jr, et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet*. 1991;48(6):1034-1050.

**3**. Naj AC, Jun G, Beecham GW, et al. Common variants at *MS4A4/MS4A6E*, *CD2AP*, *CD33* and *EPHA1* are associated with late-onset Alzheimer's disease. *Nat Genet*. 2011;43(5):436-441.

**4**. Hollingworth P, Harold D, Sims R, et al; Alzheimer's Disease Neuroimaging Initiative; CHARGE Consortium; EADI1 Consortium. Common variants at *ABCA7*, *MS4A6A/MS4A4E*, *EPHA1*, *CD33* and *CD2AP* are associated with Alzheimer's disease. *Nat Genet*. 2011;43(5):429-435.

**5**. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45(12):1452-1458.

**6**. Coppola G, Chinnathambi S, Lee JJ, et al; Alzheimer's Disease Genetics Consortium. Evidence for a role of the rare p.A152T variant in *MAPT* in increasing the risk for FTD-spectrum and Alzheimer's diseases. *Hum Mol Genet*. 2012;21(15):3500-3512.

**7**. Guerreiro R, Wojtas A, Bras J, et al; Alzheimer Genetic Analysis Group. *TREM2* variants in Alzheimer's disease. *N Engl J Med*. 2013;368(2):117-127.

**8**. Rademakers R, Neumann M, Mackenzie IR. Advances in understanding the molecular basis of frontotemporal dementia. *Nat Rev Neurol*. 2012;8(8):423-434.

**9**. Baker M, Litvan I, Houlden H, et al. Association of an extended haplotype in the *tau* gene with progressive supranuclear palsy. *Hum Mol Genet*. 1999;8(4):711-715.

**10**. Pittman AM, Myers AJ, Abou-Sleiman P, et al. Linkage disequilibrium fine mapping and haplotype association analysis of the *tau* gene in progressive supranuclear palsy and corticobasal degeneration. *J Med Genet*. 2005;42(11):837-846.

**11**. Höglinger GU, Melhem NM, Dickson DW, et al; PSP Genetics Study Group. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet*. 2011;43(7):699-705.

**12**. Van Deerlin VM, Sleiman PMA, Martinez-Lage M, et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet*. 2010;42(3):234-239.

**13**. Ferrari R, Hernandez DG, Nalls MA, et al. Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol*. 2014;13(7):686-699.

**14**. Lee SH, Harold D, Nyholt DR, et al; ANZGene Consortium, International Endogene Consortium, the Genetic and Environmental Risk for Alzheimer's Disease (GERAD1) Consortium. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet*. 2013;22(4):832-841.

**15**. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63(2):168-174.

**16**. Huyghe JR, Jackson AU, Fogarty MP, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet*. 2013;45(2):197-201.

**17**. Wang H, St Julien KR, Stevenson DK, et al. A genome-wide association study (GWAS) for bronchopulmonary dysplasia. *Pediatrics*. 2013;132(2):290-297.

**18**. Peloso GM, Auer PL, Bis JC, et al; NHLBI GO Exome Sequencing Project. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet*. 2014; 94(2):223-232.

**19**. Holmen OL, Zhang H, Zhou W, et al. No large-effect low-frequency coding variation found for myocardial infarction. *Hum Mol Genet*. 2014;23 (17):4721-4728.

**20**. Chung SJ, Kim MJ, Kim J, et al. Exome array study did not identify novel variants in Alzheimer's disease. *Neurobiol Aging*. 2014;35(8):1958.e13-1958.e14.

**21**. Coppola G, Miller BL, Chui H, et al. Genetic investigation in frontotemporal dementia and Alzheimer's disease: the GIFT Study. Ann Neurol. 2007;62(suppl 11):S52. Abstract T-10. http: //onlinelibrary.wiley.com/doi/10.1002/ana.v62:11 %2B/issuetoc. Accessed December 19, 2014.

**22**. Li Y, Chen JA, Sears RL, et al. An epigenetic signature in peripheral blood associated with the haplotype on 17q21.31, a risk factor for neurodegenerative tauopathy. *PLoS Genet*. 2014;10 (3):e1004211. doi:10.1371/journal.pgen.1004211.

**23**. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011;89(1):82-93.

**24**. Zhang B, Gaiteri C, Bodea LG, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 2013;153(3):707-720.

**25**. Webster JA, Gibbs JR, Clarke J, et al; NACC-Neuropathology Group. Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet*. 2009;84(4):445-458.

**26**. Galvin JE, Palamand D, Strider J, Milone M, Pestronk A. The muscle protein dysferlin accumulates in the Alzheimer brain. *Acta Neuropathol*. 2006;112(6):665-671.

**27**. Rademakers R, Cruts M, Sleegers K, et al. Linkage and association studies identify a novel locus for Alzheimer disease at 7q36 in a Dutch population-based sample. *Am J Hum Genet*. 2005; 77(4):643-652.

**28**. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*. 2014;133(1):1-9.

**29**. Reitz C, Jun G, Naj A, et al; Alzheimer Disease Genetics Consortium. Variants in the ATP-binding cassette transporter (*ABCA7*), apolipoprotein E ε4,and the risk of late-onset Alzheimer disease in African Americans. *JAMA*. 2013;309(14):1483-1492.

**30**. Liu J, Aoki M, Illa I, et al. Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nat Genet*. 1998;20(1):31-36.

**31**. Bashir R, Britton S, Strachan T, et al. A gene related to *Caenorhabditis elegans* spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. *Nat Genet*. 1998;20(1):37-42.

**32**. Bansal D, Miyake K, Vogel SS, et al. Defective membrane repair in dysferlin-deficient muscular dystrophy. *Nature*. 2003;423(6936):168-172.

**33**. Han R, Campbell KP. Dysferlin and muscle membrane repair. *Curr Opin Cell Biol*. 2007;19(4): 409-416.

**34**. Hochmeister S, Grundtner R, Bauer J, et al. Dysferlin is a new marker for leaky brain blood vessels in multiple sclerosis. *J Neuropathol Exp Neurol*. 2006;65(9):855-865.

**35**. Cacciottolo M, Nogalska A, D'Agostino C, Engel WK, Askanas V. Dysferlin is a newly identified binding partner of AβPP and it co-aggregates with amyloid-β42 within sporadic inclusion-body myositis (s-IBM) muscle fibers. *Acta Neuropathol*. 2013;126(5):781-783.

**36**. Jowsey PA, Doherty AJ, Rouse J. Human PTIP facilitates ATM-mediated activation of p53 and promotes cellular resistance to ionizing radiation. *J Biol Chem*. 2004;279(53):55562-55569.

**37**. Munoz IM, Jowsey PA, Toth R, Rouse J. Phospho-epitope binding by the BRCT domains of hPTIP controls multiple aspects of the cellular response to DNA damage. *Nucleic Acids Res*. 2007; 35(16):5312-5322.

**38**. Wu J, Prindle MJ, Dressler GR, Yu X. PTIP regulates 53BP1 and SMC1 at the DNA damage sites. *J Biol Chem*. 2009;284(27):18078-18084.