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 × 10^-5) and PAXIP1 (P = 2.26 × 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.

Published online February 23, 2015.

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

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, and known high-penetrance rare variants account for few cases, collectively totaling only a fraction of the estimated 58% to 79% heritability of AD. 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 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, bronchopulmonary dysplasia, and heart disease. For AD, Chung et al 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. 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 misclassified samples, genetic ancestry was also estimated by multidimensional
### Table 1. Demographic Information for the Discovery Cohort

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

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

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>European (n = 135)</th>
<th>African American (n = 271)</th>
<th>Latino (n = 50)</th>
<th>Asian (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>68 (50.4)</td>
<td>138 (50.9)</td>
<td>21 (42.0)</td>
<td>13 (54.2)</td>
</tr>
<tr>
<td>Control</td>
<td>67 (49.6)</td>
<td>133 (49.1)</td>
<td>29 (58.0)</td>
<td>11 (45.8)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>68 (50.4)</td>
<td>73 (26.9)</td>
<td>19 (38.0)</td>
<td>8 (33.3)</td>
</tr>
<tr>
<td>Female</td>
<td>57 (42.2)</td>
<td>198 (73.1)</td>
<td>31 (62.0)</td>
<td>16 (66.7)</td>
</tr>
<tr>
<td>Unknown</td>
<td>10 (7.4)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Contributing center</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emory University</td>
<td>21 (15.6)</td>
<td>223 (82.3)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>University of California, Berkeley</td>
<td>33 (24.4)</td>
<td>14 (5.2)</td>
<td>8 (16.0)</td>
<td>8 (33.3)</td>
</tr>
<tr>
<td>University of California, Davis</td>
<td>3 (2.2)</td>
<td>32 (11.8)</td>
<td>23 (46.0)</td>
<td>5 (20.8)</td>
</tr>
<tr>
<td>University of California, Irvine</td>
<td>55 (40.7)</td>
<td>2 (0.7)</td>
<td>5 (10.0)</td>
<td>1 (4.2)</td>
</tr>
<tr>
<td>University of California, Los Angeles</td>
<td>2 (1.5)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>University of California, San Francisco</td>
<td>20 (14.8)</td>
<td>0</td>
<td>0</td>
<td>6 (25.0)</td>
</tr>
<tr>
<td>University of Southern California</td>
<td>1 (0.7)</td>
<td>0</td>
<td>14 (28.0)</td>
<td>4 (16.7)</td>
</tr>
<tr>
<td>APOE genotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2/E2</td>
<td>1 (0.7)</td>
<td>2 (0.7)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E2/E3</td>
<td>4 (3.0)</td>
<td>16 (5.9)</td>
<td>2 (4.0)</td>
<td>1 (4.2)</td>
</tr>
<tr>
<td>E2/E4</td>
<td>5 (3.7)</td>
<td>9 (3.3)</td>
<td>3 (6.0)</td>
<td>0</td>
</tr>
<tr>
<td>E3/E3</td>
<td>41 (30.4)</td>
<td>87 (32.1)</td>
<td>34 (68.0)</td>
<td>8 (33.3)</td>
</tr>
<tr>
<td>E3/E4</td>
<td>21 (15.6)</td>
<td>86 (31.7)</td>
<td>9 (18.0)</td>
<td>3 (12.5)</td>
</tr>
<tr>
<td>E4/E4</td>
<td>9 (6.7)</td>
<td>12 (4.4)</td>
<td>1 (2.0)</td>
<td>2 (8.3)</td>
</tr>
<tr>
<td>Untyped</td>
<td>54 (40.0)</td>
<td>59 (21.8)</td>
<td>1 (2.0)</td>
<td>10 (41.7)</td>
</tr>
</tbody>
</table>

Abbreviation: AD, Alzheimer disease.
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, 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), 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 Hi haplotype (89.6% Hi/Hi and 10.4% Hi/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 × 10−7; minor allele odds ratio [OR], 3.0) and rs4420638 (P = 2.58 × 10−6; minor allele OR, 2.3). Both variants are within the APOE/TOMM40/APOC1 region on chromosome 19 identified in previous genetic studies.25 One variant was associated with FTD, exm2250002 (P = 2.08 × 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 × 10−3) 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

### Table 3. GCTA Explained Variance Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>AD (Variance Explained)</th>
<th>FTD (Variance Explained)</th>
<th>PSP (Variance Explained)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All exome array variants*</td>
<td>0.44 (0.39)</td>
<td>0.53 (0.36)</td>
<td>0.57 (0.44)</td>
</tr>
<tr>
<td>Exonic fraction</td>
<td>0.50 (0.36)</td>
<td>0.45 (0.35)</td>
<td>0.26 (0.56)</td>
</tr>
<tr>
<td>Low-frequency exonic fraction*</td>
<td>0.41 (0.39)</td>
<td>0.42 (0.37)</td>
<td>0.03 (0.58)</td>
</tr>
</tbody>
</table>

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

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

* Less than 5% minor allele frequency between all disease cohorts and control subjects.
The association \(-\log_{10} P\) values calculated by logistic regression are presented for Alzheimer disease, frontotemporal dementia, and progressive supranuclear palsy. The horizontal line indicates the suggestive \(P\) value threshold of \(1 \times 10^{-5}\). 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 (\(P = 2.05 \times 10^{-3}\)) 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, with 23 of 32 SNPs for which the risk allele was unambiguous showing 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, 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 17141 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.

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. DYSF 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, 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 ehm2250002 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 rare.

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 rs1417929, rs115550680, and rs3764650, as well as the missense polymorphism rs3752246. 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 muscular dystrophies such as Miyoshi myopathy and limb-girdle muscular dystrophy type 2B, known as dysferlinopathies. In skeletal muscle, dysferlin is thought to have a role in calcium-dependent sarcolemma repair. 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 and within Aβ plaques of patients with AD. 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.

The second highlighted gene, PAXIP1, encodes for a nuclear protein with 6 BRCT domains, hinting at its function in DNA repair pathways. PAXIP1 may participate in p53 activation mediated by the ataxia-telangiectasia mutated (ATM) serine/threonine kinase. Although variants in PAXIP1 have not been definitively associated with disease, Rademakers et al. 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. 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. and the linkage study by Rademakers et al. 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 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 pro-
vides one of the first glimpses at this genetic variation that hereto-
fore had not been widely studied. We anticipate that the re-
sults described herein will provide insight into the genetics of
AD, FTD, and PSP and that the data will provide a valuable mul-
tiancestral cohort with exome array genotyping data for fu-
ture studies in each of the 3 diseases. We further expect in the
long term that increased understanding of the genetic under-
pinnings will lead to improvements in diagnosis and manage-
ment for patients with neurodegenerative diseases.

REFERENCES
1. Londos CL, Ashall F, Goate AM. Exploring the etiology of Alzheimer disease using molecular genet-
Consortium; EADII Consortium. Common variants at ABCAT2, MS4A4A/MS4A4E, EPHAI, C1D3, and
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
8. Rademakers R, Neumann M, Mackenzie IR. Advances in understanding the molecular basis of
association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration.
11. Höglinger GU, Melhem NM, Dickson DW, et al; PSP Genetics Study Group. Identification of common
12. Van Deerlin VM, Sleiman PMA, Martinez-Lage M, et al. Common variants at 7q21 are associated
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

ARTICLE INFORMATION
Accepted for Publication: November 3, 2014.
Published Online: February 23, 2015.
Author Affiliations: Semel Institute for
Neuroscience and Human Behavior, University of
California, Los Angeles (Chen, Wang, Davis-Turak,
Li, Hsu, Sears, Chatzopoulos, 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 (Haeae); 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/Sponsor: This work was funded by grants
from the National Institutes of Health: F31
NS048556 (Mr Chen), PS0 AG023501 (Dr Miller),
RC1 AG035610 (Dr Coppola), RO1 MH097268 (Dr
Coppola), RO1 AG 26938 (Dr Geschwind), PO1
AG019724 (Drn Levenson, B. L. Miller, and
Geschwind), and RO1 AG10762 (Drn Levenson); from
the John Douglas French Alzheimer’s
Foundation (Drn 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 PO1 AG0542 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 Yuong, 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.

Copyright 2015 American Medical Association. All rights reserved.