

# Association of Variants in *PINX1* and *TREM2* With Late-Onset Alzheimer Disease

Giuseppe Tosto, MD, PhD; Badri Vardarajan, PhD; Sanjeev Sariya, MS; Adam M. Brickman, PhD; Howard Andrews, PhD; Jennifer J. Manly, PhD; Nicole Schupf, PhD; Dolly Reyes-Dumeyer, BS; Rafael Lantigua, MD; David A. Bennett, MD; Phillip L. De Jager, MD, PhD; Richard Mayeux, MD

**IMPORTANCE** Genetic causes of late-onset Alzheimer disease (LOAD) are not completely explained by known genetic loci. Whole-exome and whole-genome sequencing can improve the understanding of the causes of LOAD and provide initial steps required to identify potential therapeutic targets.

**OBJECTIVE** To identify the genetic loci for LOAD across different ethnic groups.

**DESIGN, SETTING, AND PARTICIPANTS** This multicenter cohort study was designed to analyze whole-exome sequencing data from a multiethnic cohort using a transethnic gene-kernel association test meta-analysis, adjusted for sex, age, and principal components, to identify genetic variants associated with LOAD. A meta-analysis was conducted on the results of 2 independent studies of whole-exome and whole-genome sequence data from individuals of European ancestry. This group of European American, African American, and Caribbean Hispanic individuals participating in an urban population-based study were the discovery cohort; the additional cohorts included affected individuals and control participants from 2 publicly available data sets. Replication was achieved using independent data sets from Caribbean Hispanic families with multiple family members affected by LOAD and the International Genetics of Alzheimer Project.

**MAIN OUTCOMES AND MEASURES** Late-onset Alzheimer disease.

**RESULTS** The discovery cohort included 3595 affected individuals, while the additional cohorts included 5931 individuals with LOAD and 5504 control participants. Of 3916 individuals in the discovery cohort, we included 3595 individuals (1397 with LOAD and 2198 cognitively healthy controls; 2451 [68.2%] women; mean [SD] age, 80.3 [6.83] years). Another 321 individuals (8.2%) were excluded because of non-LOAD diagnosis, age younger than 60 years, missing covariates, duplicate data, or genetic outlier status. Gene-based tests that compared affected individuals ( $n = 7328$ ) and control participants ( $n = 7702$ ) and included only rare and uncommon variants annotated as having moderate-high functional effect supported *PINX1* (8p23.1) as a locus with gene-wide significance ( $P = 2.81 \times 10^{-6}$ ) after meta-analysis across the 3 studies. The *PINX1* finding was replicated using data from the family-based study and the International Genetics of Alzheimer Project. Full meta-analysis of discovery and replication cohorts reached a  $P$  value of  $6.16 \times 10^{-7}$  for *PINX1* (in 7620 affected individuals vs 7768 control participants). We also identified *TREM2* in an annotation model that prioritized highly deleterious variants with a combined annotation dependent depletion greater than 20 ( $P = 1.0 \times 10^{-7}$ ).

**CONCLUSIONS AND RELEVANCE** This gene-based, transethnic approach identified *PINX1*, a gene involved in telomere integrity, and *TREM2*, a gene with a product of an immune receptor found in microglia, as associated with LOAD. Both genes have well-established roles in aging and neurodegeneration.

JAMA Neurol. doi:10.1001/jamaneurol.2019.1066  
Published online May 6, 2019.

 Supplemental content

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

**Corresponding Author:** Richard Mayeux, MD, Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, 630 W 168th St, New York, NY 10032 (rpm2@columbia.edu).

Late-onset Alzheimer disease (LOAD) is the most common form of dementia. Accumulating evidence supports a strong genetic causative mechanism. Genome-wide association studies have identified numerous loci attributed to common variants ( $\geq 5\%$  allele frequency) associated with LOAD, while next-generation sequencing has identified rare variants ( $< 5\%$ ) that may have larger effects on LOAD risk.<sup>1</sup> However, the ability to detect and confirm rare variants is limited because of low frequency. Computational approaches for assessing function and collapsing analyses have improved the ability to identify genes containing these variants.

Rare variants tend to be ethnicity specific.<sup>2,3</sup> Consequently, lack of diversity in genetic studies can be considered a potential source of bias, since LOAD frequency varies across populations,<sup>4</sup> as do underlying genetic variants. To overcome this limitation, we used whole-exome sequence (WES) data from 3595 individuals in a multiethnic, community-based study of aging and dementia, the Washington Heights, Hamilton Heights, Inwood Community Aging Project (WHICAP), as a discovery cohort. To augment the sample size for this transethnic analysis, we added sequence data from the Alzheimer Disease Sequencing Project (ADSP)<sup>5</sup> and the Religious Orders Study and Memory and Aging Project (ROSMAP).

## Methods

### Participants

The WHICAP cohort is based on a prospective, population-based study of aging and dementia including Medicare recipients residing in northern Manhattan, New York, New York. The institutional review board of the Columbia University Medical Center approved recruitment for WHICAP, and written informed consent was obtained from all participants. A total of 3595 individuals who had completed DNA and cognitive assessments were selected for WES at Columbia University.<sup>6</sup>

We also used data from 2 other independent case-control cohorts (eMethods 2 in the [Supplement](#)): the ADSP case-control cohort and ROSMAP. From the ADSP, we obtained WES data from 10 339 participants (5476 individuals with LOAD and 4863 healthy control participants) of European American ancestry from the case-control data set.<sup>7</sup> We included 455 affected individuals and 641 control participants from the ROSMAP, which is a longitudinal, epidemiologic, clinicopathologic cohort studies with whole-genome sequencing data (WGS).

### Alignment and Variant Calling

The reads obtained from the pooled WHICAP sequencing data were aligned to the human reference genome (Genome Reference Consortium GRCh37/hg19) using the Burrows Wheeler Aligner. Quality control of the data used established methods, including variant quality score recalibration and refinement of local alignment around putative indels using the Genome Analysis Toolkit. Variants were called and recalibrated

## Key Points

**Question** Can rare or uncommon coding variants confer risk of late-onset Alzheimer disease across different ethnic groups?

**Findings** Via this transethnic meta-analysis combining whole-exome and whole-genome sequencing data from 15 030 participants in 3 case-control studies, novel variants in a new locus *PINX1* and in a late-onset Alzheimer disease-associated gene *TREM2* were identified.

**Meaning** Genetic investigations across different ethnic groups in large study cohorts can improve understanding of late-onset Alzheimer disease genetic mechanisms and provide new, biologically testable hypotheses.

using multisample calling with the Genome Analysis Toolkit's UnifiedGenotyper and VariantRecalibrator modules.

### Variant Quality Control

Extensive description of the procedures can be found in eMethods 1 in the [Supplement](#). Briefly, we excluded monomorphic variants, variant quality score recalibration non-pass (filtered) variants with call rates less than 80%, and variants with low mean depth of data (less than 8 or greater than 500 reads) and genotype quality (less than 20). These thresholds were chosen to be consistent with concordance between sequencing experiments and genotyping arrays, achieving a 99% genotype likelihood.<sup>8</sup> Because these simulations were focused on single-nucleotide variants only, for indels we additionally applied the Genome Analysis Toolkit-recommended hard filters and assessed variant quality normalized by depth (QualByDepth  $> 2.0$ ) as well as strand bias (FisherStrand  $< 200.0$ ) and alternate allele position in a read test (ReadPosRankSumTest  $> 20.0$ ). We handled multiallelic sites by splitting the alternative alleles into multiple biallelic sites and normalized variants by applying parsimonious representation and left alignment. Variants showing strong departure from Hardy Weinberg equilibrium ( $P < 1.00 \times 10^{-7}$ ) in control participants were also filtered out.

### Annotation

Quality-controlled variants were annotated by Variant Effect Predictor (VEP).<sup>9</sup> This included in-silico functional prediction with combined annotation dependent depletion (CADD).<sup>10</sup>

### Replication Cohorts

To replicate findings we used (1) the ADSP family study, with WGS data from 67 families of Caribbean Hispanic ancestry from the Estudio Familiar de Influencia Genética en la Enfermedad de Alzheimer (EFIGA;  $N = 358$  participants) and (2) data from the International Genomics of Alzheimer Project (IGAP)<sup>11</sup> (eMethods 2 in the [Supplement](#)).

For expression data, we used 3 sources of data to study differential expression between individuals with LOAD and control participants for prioritized genes. First, we used neocortical transcriptome data of 364 autopsy-confirmed LOAD and normal brains (GSE15222) obtained from Myers et al.<sup>12</sup> Additionally, in the ROSMAP RNA sequencing data, 508 participants were available for analyses, as well as data from expres-

Table 1. Effective Number of Tests per Washington Heights, Hamilton Heights, Inwood Community Aging Project by Ethnic Group<sup>a</sup>

Model	Non-Hispanic White Participants		African American Participants		Caribbean Hispanic Participants		Mean Values	
	Tests, No.	<i>P</i> Value	Tests, No.	<i>P</i> Value	Tests, No.	<i>P</i> Value	Tests, No.	<i>P</i> Value
Moderate-high effect	14 075	$3.55 \times 10^{-6}$	15 745	$3.17 \times 10^{-6}$	15 713	$3.18 \times 10^{-6}$	15 178	$3.29 \times 10^{-6}$
Loss of function	1400	$3.57 \times 10^{-5}$	1578	$3.17 \times 10^{-5}$	1574	$3.18 \times 10^{-5}$	1517	$3.29 \times 10^{-5}$
Combined annotation dependent depletion score >20	9111	$5.49 \times 10^{-6}$	10 496	$4.76 \times 10^{-6}$	10 505	$4.76 \times 10^{-6}$	10 037	$4.98 \times 10^{-6}$

<sup>a</sup> The tests column reports the number of effective test to adjust for multiple testing comparison. The *P* value column reports the necessary minimum

*P* value to reach gene-wide significance (ie, a *P* value of .05 divided by the total number of tests).

sion profiles in the human prefrontal cortex that included individuals with LOAD and control participants (GSE33000)<sup>13</sup> (eMethods 3 in the Supplement).

## Statistical Analyses

### Statistical Models

We used 2 statistical models. Model 1 adjusted for sex, age (age at onset for incident cases, baseline for prevalent cases, and last observation for control participants), and principal components reflecting the population substructure within each ethnic group that remained significant after regression on the outcome (LOAD; eFigure 6 in the Supplement). Model 2 adjusted for sex, age, *APOE* ε4 allele, and principal components. For a gene-based test, we used an optimal single-nucleotide polymorphism-set (Sequence) Kernel Association Test (SKAT-O), which combined SKAT and burden tests, filtering out common variants (minor allele frequency [MAF] >0.05) and including genes with at least 2 annotated variants.

### Annotation Models

We filtered out nonfunctional variants based on annotated algorithms using VEP.<sup>9</sup> We selected 3 annotation models and assessed the agreement between them by computing Spearman coefficients between models, using *P* values. The first annotation model focused on annotations considered to have moderate-high effect: splice acceptor, splice donor, stop gain, frameshift, stop lost, start lost, or transcript amplification, inframe insertion, inframe deletion, missense variant, or protein altering. The second annotation model focused on loss-of-function classifications and made use of Loss-Of-Function Transcript Effect Estimator,<sup>14,15</sup> a VEP plugin. We filtered out those variants affecting the first and last 5% of a gene's coding sequence, because the selective constraints in terminal regions are more relaxed.<sup>16</sup> We also filtered out low-confidence loss-of-function variants: (1) splice-site variants in small introns or an intron with noncanonical splice sites; (2) stop-gained variants in the last 5% of the transcript or in an exon with noncanonical splice sites around it. In the third annotation model, combined annotation dependent depletion was used to quantitatively prioritize functional, deleterious, and disease-causing variants across a wide range of functional categories, effect sizes, and genetic architectures.<sup>10</sup> Using their Phred score implemented in a VEP plugin, we selected CADD of 20 or greater (ie, the 1% most deleterious variants).

### Bioinformatic Tools

We used the R software (R Foundation for Statistical Computing) with the package SKAT for single-marker analysis within

each the 3 WHICAP ethnic subgroups, ADSP sample, and ROSMAP sample. The SKAT Binary function was used with the SKAT-O option; we also applied the small sample adjustment available in SKAT when the size of a given group was less than 2000. Transethnic meta-analysis across the 3 WHICAP subgroups was performed using R package MetaSKAT.<sup>17</sup> We assumed that genetic effects were heterogeneous across study cohorts, and therefore we used study-specific MAFs to calculate weights. The MetaSKAT coding was used by setting a logical value for determining genetic effects of a single-nucleotide polymorphism equal to true for heterogeneous genetic effects and combined weight set equal to false to determine study-specific MAFs. Meta-analysis between the WHICAP, ADSP, and ROSMAP cohorts was an inverse variance-weighted meta-analysis based on *P* values and sample size.

Individuals from the Caribbean Hispanic families with WGS data were analyzed through the package Efficient and Parallelizable Association Container Toolbox (Center for Statistical Genetics, University of Michigan<sup>18</sup>); we conducted a gene-wise variable-threshold burden test using Efficient Mixed-Model Association Expedited (EMMAX) (Center for Statistical Genetics, University of Michigan).<sup>19</sup> We accounted for familial relationships by estimating a kinship matrix through a function implemented in the Efficient and Parallelizable Association Container Toolbox using sequencing data, filtering in only single-nucleotide variants with a high callrate (95% or higher) and an MAF of 5% or greater.

As an additional replication, we used the single-marker summary statistics from IGAP and conducted gene-based association tests using the fastBAT module in Genome-Wide Complex Trait Analysis (CNS Genomics).<sup>20</sup> Because we did not have access to imputed genetic data, we used a 1000-genome linkage disequilibrium matrix (with samples of white individuals) to estimate linkage disequilibrium between each pair of variants. We used data from Myers et al,<sup>12</sup> microarray expression data, ROSMAP RNA sequencing data, and data from Narayanan et al<sup>13</sup>; analysis methods are described in the Supplement (eMethods 3 in the Supplement).

### Multiple Testing Correction and Statistical Output

For the WHICAP analyses, we estimated the effective number of tests for Bonferroni correction using the minimum achievable *P* values in the SKAT package. Table 1 illustrates the number of effective tests per ethnic group in WHICAP and the corresponding *P* value threshold to declare gene-wide significance. Estimation of the effective number of tests using minimum achievable *P* values provides a simple and fast alterna-

Table 2. Demographics for Cohort Studies Included in the Analyses

Characteristic	WHICAP, No. (%)			ADSP, No. (%)	ROSMAP, No. (%)	EFIGA, No. (%)
	White Participants	African American Participants	Hispanic Participants			
Total samples, No.	845	1051	1699	10 339	1096	358
Individuals with late-onset Alzheimer disease	170 (20.1)	372 (35.4)	855 (50.3)	5476 (52.9)	455 (41.5)	292 (81.5)
Women	496 (58.7)	740 (70.4)	1215 (71.5)	5972 (57.7)	721 (65.8)	192 (57.6)
Age, mean (SD), y						
Total	80.7 (7)	80.1 (7)	80.1 (7)	81.0 (9)	88.1 (6)	76.9 (9)
Control participants	79.5 (7)	79.1 (7)	79.5 (7)	86.5 (4)	87.0 (6)	74.8 (8)
Affected individuals	85.4 (7)	82.1 (7)	80.8 (6)	76.1 (9)	89.7 (6)	77.2 (9)
Individuals with <i>APOE</i> ε4 allele	189 (22.4)	365 (34.7)	412 (24.2)	3055 (29.5)	279 (25.5)	96 (26.8)

Abbreviations: ADSP, The Alzheimer Disease Sequencing Project; EFIGA, Estudio Familiar de Influencia Genética en la Enfermedad de Alzheimer; ROSMAP, The Religious Orders Study and Memory and Aging Project; WHICAP, Washington Heights, Hamilton Heights, Inwood Community Aging Project.

tive to performing experimentwise permutation of the total sample to control the familywise error rate.

SKAT, a test that aggregates individual score test statistics in a single-nucleotide polymorphism-set (eg, gene or region) for association between a set of variants and dichotomous or quantitative phenotypes, computes *P* values while adjusting for covariates, such as principal components, to account for population stratification. It does not compute an effect size in the traditional sense.

### Detection of Systematic Biases

Quality control measures for the gene-based test in WHICAP meta-analysis included testing for genomic inflation factor ( $\lambda$ ) and using quantile-quantile plots. Genomic inflation factors were derived by converting *P* values to  $\chi^2$  statistics using the `qchisq` function in R, and then computing the median of this value divided by 0.456. Both tests were performed in R.

## Results

Table 2 shows the demographics stratified by ethnic group. There were 2451 women (68.2%) and 1144 men (31.8%) included. The overall mean (SD) age was 80.3 [6.83] years.

### Single Marker Analysis

No single variant reached genome-wide significance in model 1 or model 2 in WHICAP. The most significant variant was detected throughout the combined WHICAP cohort on chromosome 18, *rs116219171*, lying in the *C18orf63* gene with a  $P9.48 \times 10^{-6}$  (MAF, 0.004). The variant was not replicated in the other data sets.

### Gene-Based Analysis

In the moderate-high model, we annotated 568 348 quality controlled variants according to the VEP algorithm assigned to 18 956 genes. No significant inflation was observed after investigating the quantile-quantile plot and computing lambda value ( $\lambda = 1.05$ ). The quantile-quantile plot is shown in eFigure 2 in the Supplement. The mean number of genes to be ad-

justed for multiple testing was 15 178, with a corresponding *P* value threshold of  $3.29 \times 10^{-6}$  (defined by a Bonferroni correction of *P* value .05 by dividing by 15 178 genes; Table 1). No genes had gene-wide significance in the WHICAP meta-analysis alone; the gene with results closest to significance was *ZBTB38* on chromosome 3.

A total of 992 genes showed a SKAT-O *P* values of .05 or less in the WHICAP transethnic meta-analysis and were meta-analyzed along with ROSMAP and ADSP data (Table 3; eFigure 1 in the Supplement). On meta-analysis, we identified 1 gene that was significant after multiple-testing correction in model 1, *PINX1* ( $P = 2.81 \times 10^{-6}$ ) at 8p23.1. We also found 2 significant genes via model 2 via meta-analysis: *PINX1* ( $P = 2.10 \times 10^{-6}$ ) and *ZNF773* ( $P = 2.66 \times 10^{-6}$ ). Quantile-quantile plots for ROSMAP and ADSP are shown in eFigures 3 and 4 in the Supplement.

When each WHICAP ethnic group was tested separately, *PINX1* was found to be nominally significant in Caribbean Hispanic individuals in the WHICAP cohort ( $P = .05$ ). In the ADSP data, *PINX1* was nominally significant, with a *P* value of  $4.1 \times 10^{-4}$  in model 1 and a *P* value of  $3.5 \times 10^{-4}$  in model 2. In the ROSMAP data, SKAT-O tests resulted in a *P* value of  $9.1 \times 10^{-3}$  in model 1 and *P* value of  $6.1 \times 10^{-3}$  in model 2.

Further, *PINX1* was successfully replicated in the EFIGA families with who had WGS data (in model 1, per a EMMAX variable-threshold burden test,  $P = .04$ ; in model 2,  $P = .05$ ). The finding of significance in *ZNF773* was not replicated in either model. In addition, *PINX1* was found to be associated with LOAD in IGAP (gene-based  $P = 5.7 \times 10^{-3}$ ).

Finally, we performed an additional trans-ethnic meta-analysis including all cohorts with WES or WGS data (ie, WHICAP, ADSP, ROSMAP, and EFIGA). Through this approach, we found that *PINX1* reached a *P* value of  $8.33 \times 10^{-7}$  in model 1 and a *P* value of  $6.16 \times 10^{-7}$  in model 2, which included adjustment by *APOE* allele. Both models would be considered to have gene-wide significance after multiple-testing adjustment, according to the number of genes analyzed in the moderate-high annotation model (Table 1) but also considering the 3 annotation models presented (ie, the gene-wide *P* value threshold of  $3.29 \times 10^{-6}$  divided by 3 models results in a  $P = 1.10 \times 10^{-6}$ ).

Table 3. Top Genes From (Sequence) Kernel Association Test Meta-analysis in Each Annotation Model

Chromosome-Gene	Model 1 <sup>a</sup>				Model 2 <sup>a</sup>			
	WHICAP	ADSP	ROSMAP	Meta-analysis	WHICAP	ADSP	ROSMAP	Meta-analysis
Moderate-high effect per variant effect predictor								
19-ZNF773								
P value	6.6 × 10 <sup>-3</sup>	1.5 × 10 <sup>-4</sup>	0.55	3.57 × 10 <sup>-6</sup>	4.5 × 10 <sup>-3</sup>	2.2 × 10 <sup>-4</sup>	0.37	2.66 × 10 <sup>-6b</sup>
Variants, No.	31	49	11		31	49	11	
8-PINX1								
P value	0.031	4.1 × 10 <sup>-4</sup>	9.1 × 10 <sup>-3</sup>	2.81 × 10 <sup>-6b</sup>	0.034	3.5 × 10 <sup>-4</sup>	6.1 × 10 <sup>-3</sup>	2.10 × 10 <sup>-6b</sup>
Variants, No.	44	54	12		44	54	12	
Combined annotation dependent depletion >20								
6-TREM2								
P value	0.01	4.00 × 10 <sup>-5</sup>	0.44	1.27 × 10 <sup>-6b</sup>	9.7 × 10 <sup>-3</sup>	2 × 10 <sup>-5</sup>	0.48	7.76 × 10 <sup>-7b</sup>
Variants, No.	4	21	5		4	21	5	

Abbreviations: ADSP, The Alzheimer Disease Sequencing Project; ROSMAP, the Religious Orders Study and Memory and Aging Project; SKAT-O, (Sequence) Kernel Association Test-Optimal; WHICAP, Washington Heights, Hamilton Heights, Inwood Community Aging Project.

<sup>a</sup> Model 1 was adjusted for sex, age, and principal components. Model 2 was

adjusted for sex, age, *APOE-ε4* allele, and principal components. The stated number of variants indicates the number of single-nucleotide variants and indels included in each cohort.

<sup>b</sup> P values that passed the gene-wide significance thresholds.

### Loss-of-Function Model

We annotated 26 766 variants as loss-of-function assigned to 6406 genes. A minimum *P* value of  $3.29 \times 10^{-5}$  was necessary to define gene-wide significance. No genes reached this level of significance in the WHICAP data, which contained only 2 loss-of-function variants. While 200 genes in the WHICAP data showed SKAT-O *P* values of .05 or less and were meta-analyzed with the ADSP and ROSMAP data, none of these were found to have gene-wide significance in either model. The *P* value closest to significance found in model 1 was  $5.8 \times 10^{-4}$ , and this was with respect to *RRPI2*. In model 2, the *P* value closest to significance was found on meta-analysis was  $4.3 \times 10^{-4}$  for *ERICH6* on chromosome 3.

The CADD scores produced a wider set of variants (276 944 variants in 17 207 genes), but none had gene-wide significance. After meta-analysis across WHICAP, ADSP, and ROSMAP, *TREM2* showed gene-wide significance with a *P* value of  $1.3 \times 10^{-6}$  in model 1 and a *P* value of  $7.8 \times 10^{-7}$  in model 2.

All variants included in the analyses for *PINX1* and *TREM2* can be found in the eTable 1 and eTable 2 in the Supplement, respectively. The MAF derived from the 1000 genomes and Exome Aggregation Consortium are reported in eTable 3 in the Supplement. Another 8 LOAD-known loci reached nominal significance in the combined meta-analysis, including the *MSA4* cluster, *NME8*, and *SORL1* (eTable 4 in the Supplement). The correlation between VEP moderate-high and CADD20 models was also found to be high (eFigure 5 in the Supplement).

The *PINX1* gene was found to be significantly overexpressed in individuals whose brains showed signs consistent with LOAD vs control participants in the Myers et al<sup>12</sup> data (ratio, 1.23 [95% CI, 1.15-1.32]; *P* =  $1.06 \times 10^{-5}$ ; eFigure 7 in the Supplement) and in the Narayanan et al<sup>13</sup> data (ratio, 1.06 [95% CI, 1.04-1.08]; *P* =  $1.72 \times 10^{-10}$ ). In ROSMAP, the ratio of *PINX1* expression was similar to other data sets (ratio, 1.19; *P* = .58) and was not statistically significant. As observed in an earlier study,<sup>21</sup> *TREM2* was not significantly associated with amyloid burden or with tau burden.

### Discussion

To our knowledge, this is the largest gene-based, transethnic meta-analysis of sequencing data in LOAD to date. We identified 2 genes using different annotation models. In the moderate-high effect model, we identified *PINX1* (*PIN2*/telomeric repeat-binding factor 1 [*TERF1*]-interacting telomerase inhibitor 1). Telomerase is responsible for the maintenance of telomeres, guanine-rich noncoding tandem-repeated DNA sequences located at the ends of eukaryotic chromosomes. They preserve the coding region and maintain chromosomal integrity. Telomeres shorten after each cell cycle until reaching a critical length when the terminal segments are more likely to break. This process triggers DNA damage response machinery, leading to cellular senescence or apoptosis. The role of telomeres in aging and neurodegeneration has been extensively studied.<sup>22</sup> We previously reported that telomere length is associated with both dementia and mortality and may be interpreted as a marker of biological aging.<sup>23</sup> The *PINX1* protein differs from other proteins that regulate telomere length because it acts on telomerase, while other proteins adjust telomere length without affecting telomerase activity. Although *PINX1* interacts with telomeric repeat factor 1 and human telomerase reverse transcriptase to maintain telomere integrity, the importance and mechanisms of this physical association is not well understood.

Importantly, *PINX1* was replicated in the EFIGA family data set and in the IGAP data set, allowing us to replicate the association results in 2 different ethnic groups (Caribbean Hispanic families and a case-control group of non-Hispanic white individuals, respectively). The meta-analysis combining data from all cohorts, including those in the replication, strengthened the association, in that *PINX1* reached a *P* value of  $6.16 \times 10^{-7}$  in the *APOE*-adjusted model.

Significant overexpression of *PINX1* in 2 publicly available data sets<sup>12,13</sup> provides additional support for a putative role

for this gene in LOAD. Although this finding was not confirmed in the ROSMAP data,<sup>21</sup> the ratio of expression was similar to that in the other data sets. Of note, RNA sequencing and WES evaluate the protein-coding portion of the genome, but they approach analyses from different starting points. Gene expression varies over time, from individual to individual in different brain regions and with respect to different experiment-specific artifacts (eg, tissue quality, tissue processing).

We included rare (MAF < 1%) and uncommon variants (ie, MAF, 1%-5%) in these analyses. We justified the inclusion of common variants up to 5% because of the known variability in allele frequencies across ethnic groups (eg, a variant classified as uncommon in 1 ethnic group might be rare or ultrarare in another). The importance of uncommon variants is supported by the significant signal in *PINX1* observed in IGAP, where only variants with MAF greater than 1% were included.

We confirmed the role of *TREM2*, which was originally identified by WGS studies in which a low-frequency coding variant was found to increase the LOAD risk by approximately 2-fold to 3-fold.<sup>24-26</sup> In this study, *TREM2* was nominally significant in the WHICAP data and had gene-wide significance in the final WHICAP, ADSP, and ROSMAP meta-analysis. This confirms the established involvement of this gene in individuals with LOAD across different ethnic groups.<sup>27</sup> The original variant reported in *TREM2* was rs75932628, resulting in an arginine-to-histidine change at amino acid 47.<sup>24,25</sup> We did not identify this specific variant in WHICAP but did find 4 other variants with CADD scores of 20 or greater. One, p.H157Y (rs2234255), has already been associated with LOAD in a large meta-analysis of Han Chinese individuals,<sup>28</sup> with a large effect size (odds ratio, 3.65; MAF, 0.00103), and a recent case-control study of individuals of European descent.<sup>29</sup> These 2 studies also identified another variant also found in WHICAP, p.E151K (rs79011726) with similar conclusions. Finally, we report a novel variant, not present in the Exome Aggregation Consortium,<sup>30</sup> 6:41130779:A:G, a loss-of-function single-nucleotide variant annotated as high confidence that was significant in the WHICAP single-marker meta-analysis.

### Limitations

This study has limitations. First, most of the data came from individuals of non-Hispanic white European ancestry, whereas

the Caribbean Hispanic and African American groups were smaller in comparison. We meta-analyzed sequencing data processed from different pipelines and, most importantly, used different exome capture-methods to increase sample size. The ROSMAP study and the replication cohort of 67 Caribbean Hispanic families provided WGS. Therefore, the exome data were less comprehensive and possibly affected our ability to identify a comparable set of coding variants passing quality control, when we compared these to the WHICAP and ADSP WES studies. It is possible that missing variants in 1 or more cohorts were assumed to be monomorphic because we are not able to distinguish between missing information (eg, a variant with poor coverage, which might be excluded during the quality-control step) and monomorphic variants when variant-calling format files are meta-analyzed.

Results from this analysis that included the ADSP case-control cohort are different from what has been published.<sup>6,31</sup> First, ADSP results were derived from combining individuals with European and Caribbean Hispanic ancestries. Caribbean Hispanic individuals were not included in ADSP in this study because they were already part of the WHICAP cohort. Second, in the previous analyses of the ADSP case-control cohort,<sup>31</sup> the main statistical model accounted for population stratification but was not adjusted for sex, age, or *APOE* ε4 status. The rationale given for that approach was to maximize differences in cases and controls based on that set of covariates. These assumptions were not applied to WHICAP or ROSMAP cohorts here, and we consistently applied all potential confounders in the statistical models.

Finally, a limitation of multiple correction testing is that it applies a conservative correction in the presence of the strong correlations of variants between genes. However, gene-based tests will be less correlated than single-variant tests, because they involve multiple variants and genes located further away from each other than individual variants are.

### Conclusions

Taken together, these results indicate that *PINX1* is associated with LOAD across ethnic groups. However, the mechanisms underlying its putative role need to be established.

#### ARTICLE INFORMATION

**Accepted for Publication:** February 16, 2019.

**Published Online:** May 6, 2019.  
doi:10.1001/jamaneurol.2019.1066

**Author Affiliations:** Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University, New York, New York (Tosto, Vardarajan, Sariya, Brickman, Andrews, Manly, Schupf, Reyes-Dumeyer, Lantigua, De Jager, Mayeux); The Gertrude H. Sergievsky Center, College of Physicians and Surgeons, Columbia University, New York, New York (Tosto, Vardarajan, Sariya, Brickman, Andrews, Manly, Schupf, Reyes-Dumeyer, Lantigua, De Jager, Mayeux); Department of Neurology, College of Physicians and Surgeons, Columbia University, the New York

Presbyterian Hospital, New York (Tosto, Brickman, Manly, Schupf, De Jager, Mayeux); Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, New York (Andrews, Mayeux); Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York (Schupf, Mayeux); Department of Medicine, College of Physicians and Surgeons, Columbia University, The New York Presbyterian Hospital, New York (Lantigua); Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois (Bennett).

**Author Contributions:** Drs Mayeux and Tosto had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**Concept and design:** Tosto, Lantigua, Mayeux.  
**Acquisition, analysis, or interpretation of data:**

Tosto, Vardarajan, Sariya, Brickman, Andrews, Manly, Schupf, Reyes-Dumeyer, Bennet, De Jager, Mayeux.

**Drafting of the manuscript:** Tosto, Sariya, Andrews, Mayeux.

**Critical revision of the manuscript for important intellectual content:** Tosto, Vardarajan, Sariya, Brickman, Manly, Schupf, Reyes-Dumeyer, Lantigua, Bennet, De Jager, Mayeux.

**Statistical analysis:** Tosto, Vardarajan, Sariya, De Jager, Mayeux.

**Obtained funding:** Tosto, Manly, Bennet, Mayeux.

**Administrative, technical, or material support:** Brickman, Manly, Schupf, Reyes-Dumeyer, Bennet.

**Supervision:** Manly, Lantigua, Bennet, Mayeux.

**Conflict of Interest Disclosures:** Dr Andrews reported grants from the National Institutes of Health during the conduct of the study. Dr Manly

reported grants from the National Institute on Aging/National Institutes of Health during the conduct of the study. Dr Schupf reported grants from the National Institute on Aging/National Institutes of Health during the conduct of the study. Dr Bennet reported grants from the National Institutes of Health during the conduct of the study. No other disclosures were reported.

**Funding/Support:** This study was supported by the National Institute on Aging (grants U01 AG032984, UFIAG047133, R01 AG033193, U01AG049505, U01AG049506, U01AG049507, U01AG049508, U01AG052411, U01AG052410, U01 AG052409, U54AG052427, U01AG00678, RFIAG054023, RFIAG015473, R01AG048927, RFIAG057519, and R03AG054936), the National Library of Medicine (grant R01LM012535), the National Center for Advancing Translational Sciences (grant TL1TR001875), the National Institute of Neurological Disorders and Stroke (grant R01 NS017950), the National Heart, Lung, and Blood Institute, and the National Institutes of Health (grant UFIAG047133). The Religious Orders Study and Rush Memory and Aging Project are supported by the National Institute on Aging (grants P30AG10161, R01AG15819, R01AG17917, R0136836, and U01AG46152). Biogen Inc provided support for whole-exome sequencing for the Washington Heights, Hamilton Heights, and Inwood Community Aging Project cohort.

**Role of the Funder/Sponsor:** The funders 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.

**Disclaimer:** The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US Department of Health and Human Services.

**Additional Contributions:** We acknowledge the Washington Heights, Hamilton Heights, Inwood Community Aging Project study participants and the research and support staff for their contributions to this study. David Goldstein, PhD, the Institute of Genomic Medicine at Columbia University Medical Center, provided advice for the statistical analyses. These individuals were not compensated for their contributions.

## REFERENCES

- Tosto G, Reitz C. Genomics of Alzheimer's disease: value of high-throughput genomic technologies to dissect its etiology. *Mol Cell Probes*. 2016;30(6):397-403. doi:10.1016/j.mcp.2016.09.001
- Gravel S, Henn BM, Gutenkunst RN, et al; 1000 Genomes Project. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A*. 2011;108(29):11983-11988. doi:10.1073/pnas.1019276108
- Igartua C, Myers RA, Mathias RA, et al. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. *Nat Commun*. 2015;6:5965. doi:10.1038/ncomms6965
- Tang M-X, Stern Y, Marder K, et al. The APOE-ε4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *JAMA*. 1998;279(10):751-755. doi:10.1001/jama.279.10.751
- Beecham GW, Bis JC, Martin ER, et al. The Alzheimer's Disease Sequencing Project: study design and sample selection. *Neurol Genet*. 2017;3(5):e194. doi:10.1212/NXG.0000000000000194
- Raghavan NS, Brickman AM, Andrews H, et al; Alzheimer's Disease Sequencing Project. Whole-exome sequencing in 20,197 persons for rare variants in Alzheimer's disease. [correction published in *Ann Clin Transl Neurol*. 2019;6(2):416]. *Ann Clin Transl Neurol*. 2018;5(7):832-842. doi:10.1002/acn3.582
- The Alzheimer's Disease Sequencing Project. Welcome to the Alzheimer's Disease Sequencing Project. <https://www.niagads.org/adsp/content/home>. Accessed April 3, 2019.
- Carson AR, Smith EN, Matsui H, et al. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics*. 2014;15:125. doi:10.1186/1471-2105-15-125
- McLaren W, Gil L, Hunt SE, et al. The ensemble variant effect predictor. *Genome Biol*. 2016;17(1):122. doi:10.1186/s13059-016-0974-4
- Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315. doi:10.1038/ng.2892
- 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. doi:10.1038/ng.801
- Myers AJ, Gibbs JR, Webster JA, et al. A survey of genetic human cortical gene expression. *Nat Genet*. 2007;39(12):1494-1499. doi:10.1038/ng.2007.16
- Narayanan M, Huynh JL, Wang K, et al. Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. *Mol Syst Biol*. 2014;10:743. doi:10.1525/msb.20145304
- MacArthur DG, Balasubramanian S, Frankish A, et al; 1000 Genomes Project Consortium. A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 2012;335(6070):823-828. doi:10.1126/science.1215040
- Karczewski K. LOFTEE (loss-of-function transcript effect estimator). <https://github.com/konradjk/loftee>. Accessed April 3, 2019.
- Wetterbom A, Sevov M, Cavellier L, Bergström TF. Comparative genomic analysis of human and chimpanzee indicates a key role for indels in primate evolution. *J Mol Evol*. 2006;63(5):682-690. doi:10.1007/s00239-006-0045-7
- Mensah-Ablorh A, Lindstrom S, Haiman CA, et al. Meta-analysis of rare variant association tests in multiethnic populations. *Genet Epidemiol*. 2016;40(1):57-65. doi:10.1002/gepi.21939
- Center for Statistical Genomics. EPACTS. <https://genome.sph.umich.edu/wiki/EPACTS>. Accessed April 3, 2019.
- Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012;44(7):821-824. doi:10.1038/ng.2310
- Bakshi A, Zhu Z, Vinkhuyzen AA, et al. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Sci Rep*. 2016;6:32894. doi:10.1038/srep32894
- Mostafavi S, Gaiteri C, Sullivan SE, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci*. 2018;21(6):811-819. doi:10.1038/s41593-018-0154-9
- Cai Z, Yan L-J, Ratka A. Telomere shortening and Alzheimer's disease. *Neuromolecular Med*. 2013;15(1):25-48. doi:10.1007/s12017-012-8207-9
- Honig LS, Schupf N, Lee JH, Tang MX, Mayeux R. Shorter telomeres are associated with mortality in those with APOE ε4 and dementia. *Ann Neurol*. 2006;60(2):181-187. doi:10.1002/ana.20894
- 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. doi:10.1056/NEJMoa1211851
- Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med*. 2013;368(2):107-116. doi:10.1056/NEJMoa1211103
- Tosto G, Reitz C. TREM2 variants and Alzheimer's disease. *Future Neurol*. 2013;8:407-410. doi:10.2217/fnl.13.22
- 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. doi:10.1001/jama.2013.2973
- Sirkis DW, Bonham LW, Aparicio RE, et al. Rare TREM2 variants associated with Alzheimer's disease display reduced cell surface expression. *Acta Neuropathol Commun*. 2016;4(1):98. doi:10.1186/s40478-016-0367-7
- Jin SC, Benitez BA, Karch CM, et al. Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum Mol Genet*. 2014;23(21):5838-5846. doi:10.1093/hmg/ddu277
- Karczewski KJ, Weisburd B, Thomas B, et al; The Exome Aggregation Consortium. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 2017;45(D1):D840-D845. doi:10.1093/nar/gkw971
- Bis JC, Jian X, Kunkle BW, et al; Alzheimer's Disease Sequencing Project. Whole exome sequencing study identifies novel rare and common Alzheimer's-associated variants involved in immune response and transcriptional regulation [published online August 14, 2018]. *Mol Psychiatry*. doi:10.1038/s41380-018-0112-7