ANNALS of Clinical and Translational Neurology
Open Access
ANA American Neurological Association since 1875

RESEARCH ARTICLE

F-box/LRR-repeat protein 7 is genetically associated with Alzheimer's disease

Giuseppe Tosto[1,2,a], Hongjun Fu[1,a], Badri N. Vardarajan[1,2], Joseph H. Lee[1,2], Rong Cheng[1,2], Dolly Reyes-Dumeyer[1,2], Rafael Lantigua[3], Martin Medrano[4], Ivonne Z. Jimenez-Velazquez[5], Mitchell S. V. Elkind[2,6], Clinton B. Wright[7], Ralph L. Sacco[7,8], Margaret Pericak-Vance[8,9], Lindsay Farrer[10], Ekaterina Rogaeva[11], Peter St George-Hyslop[11,12], Christiane Reitz[1,2,6] & Richard Mayeux[1,2,6,13,14]

[1]The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, School of Public Health, Columbia University, New York, New York
[2]The Gertrude H. Sergievsky Center, School of Public Health, Columbia University, New York, New York
[3]Department of Medicine College of Physicians and Surgeons, School of Public Health, Columbia University, New York, New York
[4]School of Medicine, Pontificia Universidad Catolica Madre y Maestra, Santiago, Dominican Republic
[5]Department of Medicine, Geriatrics Program, School of Medicine, University of Puerto Rico, San Juan, Puerto Rico
[6]Department of Neurology, Columbia University, New York, New York
[7]Evelyn F. McKnight Brain Institute and Departments of Neurology, Public Health Sciences, John T. McDonald Department of Human Genetics, University of Miami, Miami, Florida 33136
[8]Neuroscience Program, Leonard M. Miller School of Medicine, University of Miami, Miami, Florida 33136
[9]The John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida 33136
[10]Departments of Medicine (Biomedical Genetics), Neurology, Ophthalmology, Biostatistics and Epidemiology, Boston University Schools of Medicine and Public Health, Boston, Massachusetts
[11]Tanz Centre for Research in Neurodegenerative Diseases, and Department of Medicine, University of Toronto, Krembil Discovery Tower, 60 Leonard Avenue, Toronto, Ontario, Canada, M5T 2S8
[12]Department of Clinical Neurosciences, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, United Kingdom
[13]Department of Psychiatry, School of Public Health, Columbia University, New York, New York
[14]Department of Epidemiology, School of Public Health, Columbia University, New York, New York

Correspondence
Richard Mayeux, Sergievsky Center, 630 West 168th Street, Columbia University, New York, NY 10032. Tel: 212-305-2391; Fax: 212-305-2518; E-mail: rpm2@cumc.columbia.edu

Funding Information
This work was supported by federal grants from the National Institutes of Health and the National Institute on Aging, specifically P50 AG08702, R37 AG15473, RO1 AG037212; the Marilyn and Henry Taub Foundation and the Charles S Robertson Gift from the Banbury Fund (R. M.); R37 NS 29993 and the McKnight Brain Research Foundation (R. S. and M. E.); Canadian Institutes of Health Research (E. R., P. S. H.); Welcome Trust, Medical Research Council, Ontario Research Fund, and Alzheimer Society of Ontario (P. S. H.).

Received: 6 March 2015; Revised: 5 May 2015; Accepted: 14 May 2015

Annals of Clinical and Translational Neurology **2015; 2(8): 810–820**

Abstract

Objective: In the context of late-onset Alzheimer's disease (LOAD) over 20 genes have been identified but, aside *APOE*, all show small effect sizes, leaving a large part of the genetic component unexplained. Admixed populations, such as Caribbean Hispanics, can provide a valuable contribution because of their unique genetic profile and higher incidence of the disease. We aimed to identify novel loci associated with LOAD. **Methods**: About 4514 unrelated Caribbean Hispanics (2451 cases and 2063 controls) were selected for genome-wide association analysis. Significant loci were further tested in the expanded cohort that also included related family members (n = 5300). Two AD-like transgenic mice models (J20 and rTg4510) were used to study gene expression. Independent data sets of non-Hispanic Whites and African Americans were used to further validate findings, along with publicly available brain expression data sets. **Results**: A novel locus, rs75002042 in *FBXL7* (5p15.1), was found genome-wide significant in the case–control cohort (odd ratio [OR] = 0.61, P = 6.19E-09) and confirmed in the related members cohorts (OR = 0.63, P = 4.7E-08). Fbxl7 protein was overexpressed in both AD-like transgenic mice compared to wild-type littermates. Publicly available microarray studies also showed significant overexpression of Fbxl7 in LOAD brains compared to nondemented controls. single-nucleotide polymorphism (SNP) rs75002042 was in complete linkage disequilibrium with other variants in two independent non-Hispanic White and African American data sets ($0.0005 < P < 0.02$) used for replication. **Interpretation**: FBXL7, encodes a subcellular protein involved in phosphorylation-

doi: 10.1002/acn3.223

[a]These authors contributed equally to the manuscript.

dependent ubiquitination processes and displays proapoptotic activity. F-box proteins also modulate inflammation and innate immunity, which may be important in LOAD pathogenesis. Further investigations are needed to validate and understand its role in this and other populations.

Introduction

Late-Onset Alzheimer's disease (LOAD), a progressive neurodegenerative disorder, is the leading cause of dementia in the elderly.[1] The most common genetic risk factor is the *APOE-ε4* allele with an attributable risk of LOAD of 10–15%.[2] The International Genomics of Alzheimer's Project (IGAP) replicated eight known LOAD-susceptibility loci (*CLU*, *PICALM*, *CR1*, and *BIN1*, *MS4A4A*/*MS4A4E*/*MS4A6E* cluster, *ABCA7*, *CD2AP*, and *EPHA1*) found in previous genome-wide association studies (GWAS)[3,4] and described 12 novel loci.[5] However, the effect sizes associated with these single-nucleotide polymorphisms (SNPs) are small (odd ratios [OR] between 1.1 and 1.3), suggesting that a large part of the genetic component of LOAD is still unexplained.[6]

Like African Americans, Caribbean Hispanics have a unique genetic background and higher incidence rate of disease,[7] which may facilitate gene discovery in complex diseases such as LOAD. In fact, in an African-American GWAS[8] *ABCA7* was identified as a susceptibility locus with an effect size similar to that of *APOE*. Hispanics are one of the fastest growing minority groups in the U.S. with Caribbean Hispanics making up to 30% of this population. Caribbean Hispanics from the Dominican Republic represent a population isolate with a limited number of founders, evidence of inbreeding[9] and large families multiply affected by LOAD.[10] We and others have demonstrated that genetic and nongenetic risk factors have different effects on LOAD risk as compared to other ethnic groups.[11,12]

Methods

Subjects

The data analyzed were derived from three studies that had previously recruited individuals of Caribbean Hispanic ancestry: (1) the Washington Heights and Inwood Columbia Aging Project (WHICAP study[12]); (2) Estudio Familiar de Influencia Genetica en Alzheimer (EFIGA study[10]) family study; (3) Northern Manhattan Study (NOMAS study[13]). Full description of the data sets is provided in the Data S1.

Written informed consent was obtained from all participants. Recruitment of subjects in the WHICAP, EFIGA, and NOMAS studies was approved by the Institutional Review Board of the Columbia University Medical Center. The NOMAS study was also approved by the University of Miami Institutional Review Board. The study was conducted according to the principles expressed in the Declaration of Helsinki.

Phenotypes

For each data set, the diagnoses of "probable" or "possible" LOAD were defined based on the NINCDS-ADRDA criteria.[14] Participants were classified as "controls" if they were found to have no evidence of cognitive impairment or dementia diagnosis at last visit available. Individuals were excluded if they had undefined phenotype, missing covariates, known LOAD mutations or if the last age of evaluation for controls or age at onset for cases was less than 60 years.

Genotyping

Four genotyping platforms were employed and their characteristics are reported in Table S1.

Imputation

Imputation of allele dosages used the March 2012 reference panel from 1000 Genomes (1000G) – build hg37 and the IMPUTE2 software (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html[15]), applying strict variant position and strand alignment controls, prephasing and preimputation filtering. The reference panel is a multiethnic panel (with four identified super-populations, that is, Europeans [$n = 379$], Africans [$n = 246$], Asian [$n = 286$], and Admixed Americans [$n = 181$]) and has been shown to perform imputation of genotypes with high accuracy in admixed samples.[16] Because a different chip was used for genotyping each cohort, imputation was performed separately (additional information provided in the Data S1). Only imputed SNP dosages with an imputation quality estimate of $R^2 \geq 0.50$ were included in the final SNP set for the present analysis.

Power

We used CaTS – Power Calculator for GWAS[17] to compute the detectible effect size for our study under varying allele frequencies in order to simulate several potential scenarios.

Relatedness and unrelated case–control sample selection

Identity by descent (IBD) was computed using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) in order to (1) identify (latent) relatedness among subjects (2) identify duplicates and (3) select a subsample of unrelated individuals. IBD is calculated by estimating the probability of sharing 0, 1, or 2 alleles for any pair of individuals ($\pi = P\ [\text{IBD} = 2] + 0.5 \times P\ [\text{IBD} = 1]$, where P indicates probability). One participant from each duplicate pair ($\pi > 0.95$) or relative pair ($0.4 \leq \pi \leq 0.95$) was included in the unrelated sample used for association analyses, prioritizing based on nonmissing disease status and/or covariates and then higher SNP call rate.

Population stratification

Population stratification was investigated using EIGENSTRAT (EIGENSOFT 3.0, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm).[18] Tracy–Widom statistics in the software allows for determination of the number of significant principal components (PC) necessary to control for population stratification in the association analysis.[19] However, overestimation of significant PC for admixed data sets or highly unbalanced case–control samples might bias the result. In order to account for these limitations, we applied the method described by Shriner[20] that capitalizes on the smaller bias and variance of Velicer's minimum average partial test to determine the optimal number of PCs to retain.

In order to estimate PCs for all study participants (related and unrelated), we first used a maximal set of unrelated individuals to obtain SNP eigenvectors and then sample eigenvectors for the remaining related persons were computed; this approach was described by Zhu et al.[21] and implemented in the KING software.[22]

Association analysis

Associations between LOAD and SNPs passing quality control were reassessed using multivariate logistic regression in PLINK version 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) for the case–control data set. All analyses were performed with an additive genetic model (i.e., genotyped SNPs were coded 0, 1, or 2 based on the number of reference alleles while imputed SNPs based on the posterior probability of the reference allele). The primary association analyses were adjusted for sex, age, population stratification, and cohort (Model A) and sex, age, population stratification, cohort, and *APOE-ε4* (Model B). A threshold *P*-value of 5.0E-08 was set for genome-wide significance after Bonferroni correction for multiple testing.

For loci in which variants reached genome-wide significance, as well as loci previously found to be associated with LOAD in the IGAP meta-analysis, we conducted gene-based association tests using the GATES procedure.[23] We included only high-quality imputed SNPs ($R^2 \geq 0.8$) and added a 50 kb flanking region at either end of each locus to cover potential regulatory regions. We chose not to employ methods using reference panels to define linkage disequilibrium (LD) across markers because of the unique LD pattern of our population.

Systematic biases detection

Quality control measures for the association testing in the case–control sample included genomic inflation factor (lambda) and quantile–quantile (Q–Q) plot to compare the genome-wide distribution of the derived test statistic with the expected null distribution. Both tests were performed with R (http://www.r-project.org).

Expanded cohort and general estimating equation

In order to verify findings obtained in the case–control GWAS in the whole sample (related and unrelated individuals) we fitted a logistic regression model via Generalized Estimation Equation (GEE) implemented in the GWAF package in R (http://cran.r-project.org/) to test association between the affection status and imputed SNPs under additive genetic model. Each family was treated as a cluster, with independence working correlation matrix used in the robust variance estimator.

Ancestry estimation

Population ancestry was estimated using the ADMIXTURE software[24] through both unsupervised (i.e., without reference populations) and supervised admixture analyses. Full descriptions of the methods applied are reported in the Data S1.

Replication data sets

Genome-wide significant variants were compared to association results in the Caucasian Alzheimer Disease Genetics Consortium – White, non-Hispanic data set[25] and African-American data set.[8]

Ethnic variation in the LD pattern

LD strength across different ethnic studies was investigated through PLINK, the Haploview software (www.broadinstitute.org) and the varLD software.[26] The latter

quantifies the inter-sample variation in LD thus underpinning regions of different haplotypic patterns across populations through 10,000 Monte Carlo (MC) iterations.

Animal experiments

Animals

B6.Cg-Tg(PDGFB-APPSwInd) (J20) mice[27] were obtained from Dr. Lennart Mucke, and C57BL/6J mice were from the Jackson Laboratory (Bar Harbor, ME, USA). The rTg4510 mice and littermate controls were generated by crossing FVB-Tg(tetO-TauP301L)4510 ("hTau")[28] and 129S6-Tg(CaMKIIa-tTA) ("tTA")[29] breeders obtained from the Mayo Clinic (Jacksonville, FL). J20 mice were compared to littermate controls without transgene. The rTg4510 mice were compared to littermate controls containing only the tTA transgene or neither transgene. The genotypes of transgenic animals were confirmed by polymerase chain reaction. All animals are maintained in the animal facility at the Columbia University Medical Center on a 12 h light/dark cycle with food and water provided ad libitum. All animal experiments were performed in accordance with national guidelines (National Institutes of Health) and approved by the Institutional Animal Care and Use Committee of Columbia University.

Western blot assay

Animals were sacrificed by decapitation, the brains were extracted, and the cortex and hippocampus were dissected from the brain. Tissue was homogenized using a Polytron in 1x Radioimmunoprecipitation assay buffer (Thermo Scientific, Rockford, IL, USA) with protease inhibitor and phosphatase inhibitor cocktails (Sigma-Aldrich, Saint Louis, MO, USA), and the lysate was sonicated using a sonicator dismembrator (Fisher Scientific, Springfield, NJ, USA). Proteins (10 μg) were separated electrophoretically on 4–12% Bis-Tris precast polyacrylamide gels (Life Technologies, Carlsbad, CA, USA) and blotted onto nitrocellulose blotting membranes (0.2 μm, GE Healthcare, Piscataway, NJ, USA). The nonspecific binding was blocked by 1-h incubation in 1x Phosphate-buffered saline containing 0.05% Tween 20, and 5% nonfat milk. Blots were probed with mouse primary antibodies for APP/Aβ (6E10) (Covance, Princeton, NJ, USA, 1:1000), Tau (CP27) (courtesy of Peter Davies, 1:1000), FBXL7 (Santa Cruz Biotechnology, Santa Cruz, CA, USA, 1:1000), or ACTB (Sigma-Aldrich, Saint Louis, MO, USA, 1:10,000). After washing and incubation with secondary horseradish peroxidase-conjugated antibodies (Jackson ImmunoResearch, West Grove, PA, USA), membranes were developed with ECL (Immobilon™ Western Chemiluminescent HRP Substrate, EMD Millipore, Billerica, MA, USA), and digitalized images were taken using Fujifilm LAS-3000 Imager (Fujifilm, Valhalla, NY, USA). Integrated density of protein bands was analyzed using ImageJ (U. S. National Institutes of Health, Bethesda, MD, USA), and the integrated density of FBXL7 protein was normalized to the housekeeping protein ACTB and expressed as the percentage of Control. The quantitative data of western blot assay were expressed as mean $\pm$ standard errors. Significance was assessed with Student's *t*-test in R (http://cran.r-project.org).

Publicly available human brain expression data sets

We extracted findings for *Fbxl7* from two publicly expression gene expression microarray studies[30,31] previously described in detail elsewhere. A full description of the data and methods is provided in the Data S1.

Results

Sample characteristics

The clinical characteristics of each data sets are summarized in Table 1. After quality control, the total number of individuals included was 5300 (WHICAP study $n = 1803$; EFIGA study $n = 3131$; NOMAS study $n = 366$); the unrelated sample comprised 4514 persons.

Power

The minimum detectable OR was estimated to be 1.4 (inverse 0.71), based on our sample size and a $\beta = 80\%$ and assuming a minor allele frequency (MAF) of 10%, a disease prevalence of 20%, an additive genetic model and a significance level of $\alpha = 5E\text{-}08$. For smaller MAF such as 6.5% and 4.5%, minimum detectable OR increases to 1.5 and 1.6, respectively.

Population structure

PCs were computed for the unrelated sample and the first three PCs were retained in order to account for population stratification in the analysis. Individuals were plotted along with the 1000G project's reference superpopulations in order to identify potential clustering in terms of ancestry. This analysis showed most of the Caribbean Hispanic subjects distributed along the first PC, thus indicating an admixture of African and European ancestry, with a second major axis of variation suggesting Native American ancestry because of its projection along the American Admixed population (Fig. 1).

Table 1. Sample characteristics.

	Unrelated sample			Related sample[1]		
	Affected	Unaffected	Overall	Affected	Unaffected	Overall
Subjects (*n*)	2451	2063	4514	3001	2299	5300
Females (%)	66.9	66.7	66.8	66.7	66.7	66.7
Mean age in years (SD)	78.6 (8)	73.5 (8)	76.3 (8)	78.9 (8)	73.4 (8)	76.5 (8)
APOE-ε4 alleles (%)						
0	1492 (61)	1531 (74)	3023	1751 (58)	1682 (73)	3433
1	814 (33)	485 (24)	1299	1047 (35)	567 (25)	1614
2	144 (6)	37 (2)	181	202 (7)	39 (2)	241
Missing APOE (*n*)	1	10	11	1	11	12

n, number; SD, standard deviation.
[1]The related sample includes the individuals from the unrelated sample.

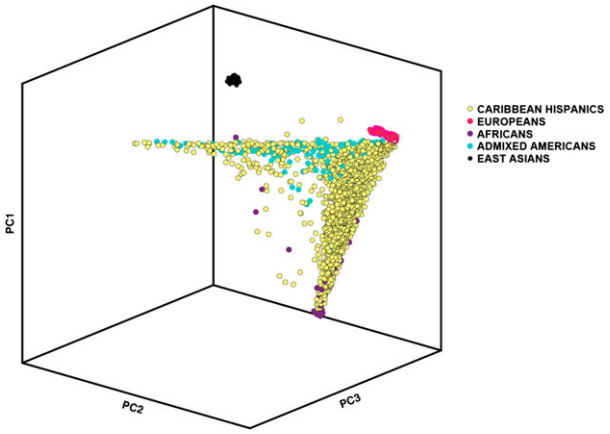

Figure 1. Top three principal components computed for the Caribbean Hispanic sample (highlighted in yellow) along with the 1000G reference superpopulations (Magenta = Europeans; Purple = African; Turquoise = American Admixed; Black = East Asians). PC1/2/3 = first, second, third principal component.

Systematic biases detection

Genomic inflation for the GWAS was minimal ($\lambda < 1.05$), therefore we did not use genomic control. Quantile-quantile plot is reported in Figure S1.

GWAS – Model A

In addition to variants lying in the *APOE* region (top SNP rs394819, OR = 0.53, SE = 0.10, P = 8.4E-11), two regions were genome-wide significant: common variants within the *FBLX7* gene on chromosome 5 (NCBI Entrez Gene 23194, 5p15.1) and *CACNA2D3* on chromosome 3 (NCBI Entrez Gene 55799, 3p21.1) with top SNPs conferring an OR of 0.61 (rs75002042, P = 6.19E-09) and 1.59 (rs7431992, P = 1.99E-08), respectively (Table 2). We generated graphic representations of the GWAS results (Manhattan plot, Fig. S2) and regional association plots for the locus on chromosome 5p15.1 (Fig. S3).

GWAS – Model B

In the fully adjusted model (sex, age, batch effect, PCs and *APOE*) the association with LOAD was confirmed, showing OR = 0.60 (P = 4.7E-09) and 1.61 (P = 5.8E-09) for SNPs rs75002042 and rs7431992, respectively.

Replication of prior published GWAS results

We investigated the 21 variants reported in the IGAP[5] meta-analysis for LOAD (nine known and 12 novel) presented in Table 3. Four of these SNPs were replicated in the current GWAS (Model A) with nominally significant *P*-values: rs17125944 in *FERMT2* (C-allele: OR = 1.22, P = 0.029), rs3865444 in *CD33* (A-allele: OR = 0.87, P = 0.008), rs10838725 in *CELF1* (C-allele: OR = 1.14, P = 0.019), and rs10498633 in the *SLC24A4-RIN3* region (T-allele: OR = 0.88, P = 0.045). Gene-based analyses further confirmed two of those loci plus an additional one: *SLC24A4-RIN3* (P = 0.01), *CD33* (P = 0.04), and *ABCA7* (P = 0.02). We also analyzed *FRMD4A*[32] in a gene-based test confirming the association in our data set (P = 0.002). Three additional loci reported by the IGAP meta-analysis (*MS4A6A*, *SORL1*, and the *HLA* region) exhibited association at a trend level with LOAD in gene-based analyses ($0.05 < P < 0.09$; data not shown).

Expanded cohort and general estimating equation

After including the related family members, the association between LOAD and SNP rs75002042 on chromosome 5p15.1 was confirmed (GEE: OR = 0.63, CI = 0.54–0.75, P = 4.7-E08) in the fully adjusted model, whereas rs7431992 on chromosome 3 showed a weaker association (GEE: OR = 1.46, CI = 1.25–1.71, P = 2.0E-06).

Table 2. Genome-wide results of main adjusted model (Model A: sex, age, cohort, principal components) and fully adjusted model (Model B: sex, age, cohort, principal components, and *APOE*): best SNPs with $P \leq 5E\text{-}08$.

CHR	Gene	SNP	BP	Minor allele	Major allele	MAF	IQ	OR	CI	*P*
Model A										
19	*TOMM40 – APOE region*	rs394819	44,901,322	T	G	0.07	0.99	1.89	1.55–2.30	8.4E-11
5	*FBXL7*	rs75002042	15,669,967	A	T	0.08	0.98	0.61	0.52–0.71	6.19E-09
3	*CACNA2D3*	rs7431992	54,353,240	A	T	0.10	0.96	1.59	1.36–1.86	1.99E-08
Model B										
5	*FBXL7*	rs75002042	15,669,967	A	T	0.08	0.98	0.60	0.51–0.70	4.7E-09
3	*CACNA2D3*	rs7431992	54,353,240	A	T	0.10	0.96	1.61	1.38–1.88	5.8E-09

Frequencies have been rounded to the second decimal. Base pair based on hg19 assembly. CHR, chromosome; SNP, single-nucleotide polymorphism; BP, base pair location; MAF, minor allele frequency; IQ, imputation quality; OR, odd ratio.

Table 3. Genome-wide significant SNPs reported in the recent IGAP meta-analysis in non-Hispanic Whites: genes are listed in the table only if found associated with LOAD in the single-marker analysis and/or gene-based analysis in the Hispanic GWAS.

									SMA		GBA	
CHR	Gene	SNP	Minor allele	Major allele	MAF	OR	CI	IQ	P^1	P^2	P^1	P^2
10	*FRMD4A*[3]	–	–	–	–	–	–	–	–	–	0.002	0.004
11	*CELF1*	rs10838725	C	T	0.20	1.14	1.01–1.28	0.98	0.019	0.010	ns	ns
14	*FERMT2*	rs17125944	C	T	0.07	1.22	1.02–1.45	0.99	0.029	0.036	ns	ns
14	*SLC24A4-RIN3*	rs10498633	T	G	0.17	0.88	0.78–0.99	1	0.045	0.043	0.01	0.04
19	*ABCA7*	rs4147929[4]	–	–	–	–	–	–	–	–	0.02	0.05
19	*CD33*	rs3865444	A	C	0.25	0.87	0.79–0.96	0.99	0.008	0.015	0.04	0.09

CHR, chromosome; SNP, single-nucleotide polymorphism; MAF, minor allele frequency; OR, odd ratio; IQ, imputation quality; SMA, single-marker analysis; GBA, gene-based analysis.

[1]Adjusted for sex, age, batch effect, PCs.

[2]Adjusted for sex, age, batch effect, PCs, and *APOE*.

[3]No SNP reported in the IGAP meta-analysis; the gene was found associated with LOAD in a haplotype GWAS.

[4]SNP reported in the IGAP meta-analysis was not present in the Hispanic GWAS.

Admixture

Supervised admixture showed that the European lineage accounted for the most part (57%), followed by African (33%) and Native American ancestry (8%). Analyses on the whole sample gained overlapping results. Results are reported in the Data S1 and illustrated in the Figure S4.

Replication data sets

SNPs lying within *FBXL7* were investigated in the publically available Alzheimer's Disease Genetics Consortium-White[25] and African-American GWAS studies.[8] As expected allele frequencies differed. SNP rs75002042 in *FBXL7* had a global allele frequency for the A-allele of 0.21 in the AA data set and 0.01 in ADGC-Whites with no association with LOAD. Because GWAS loci represent regions, not specific genes, we investigated several variants within the *FBXL7* gene that did show significant associations with the LOAD in both data sets. In the AA study, rs113637289 (OR = 1.66, CI = 1.25–2.20, $P = 4.8E\text{-}04$ in the fully adjusted model) was in complete LD with the flanking SNP rs75002042 in *FBXL7* ($D' = 1$). In the ADGC-Whites, another SNP (rs79267806) was also in complete LD ($D' = 1$) and nominally significant (OR = 1.12, CI = 1.01–1.24, $P = 0.026$). Allele frequencies for the Caribbean Hispanic cohort and the 1000G reference superpopulations are reported in Table S2.

Ethnic variation in the LD pattern

We investigated the *FBXL7* region in the 1000G project European and African superpopulations. varLD analyses confirmed strong inter-population LD variations for both loci on chromosome 5 (MC $P < 1E\text{-}04$).

Expression of Fbxl7 protein in two AD-like mouse models

We quantified the expression of Fbxl7 protein in two transgenic mice models: J20 and rTg4510. Using western blot assay, Fbxl7 protein level in the cortex and hippocampus of 10-month-old J20 mice was found increased by 95% (*t*-test, $P = 0.017$) as compared to age-matched

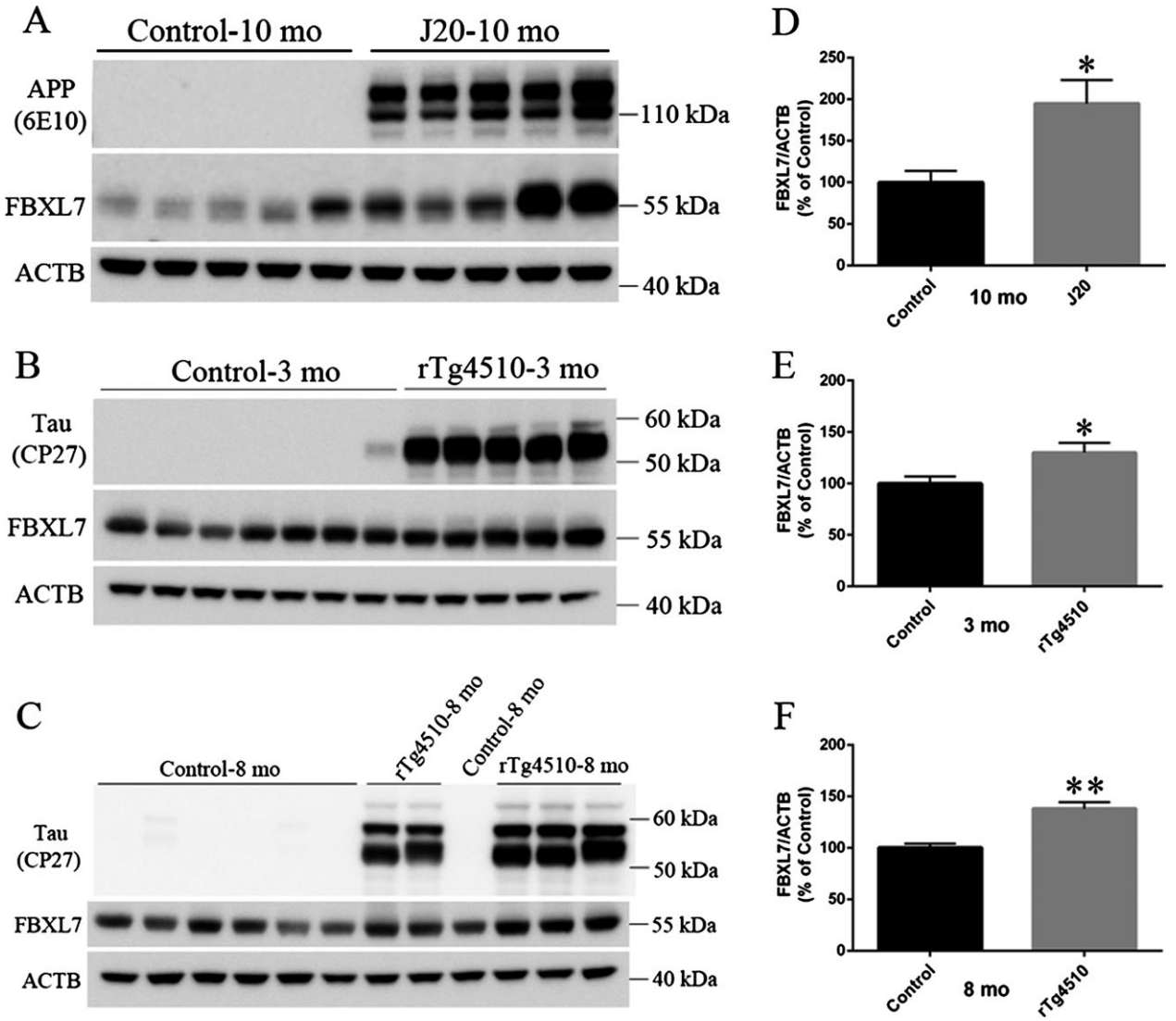

Figure 2. The expression of Fbxl7 protein in J20 and rTg4510 mice. Protein samples from the cortex and hippocampus of: (A) 10-months-old J20 (n = 5) and control mice (n = 5), (B) 3-month-old rTg4510 (n = 5) and control mice (n = 7), and (C) 8-month-old rTg4510 (n = 2) and control mice (n = 4) were separated in 4–12% Bis-Tris polyacrylamide gels and blotted with mouse primary antibodies against APP/Aβ (6E10), Tau (CP27), Fbxl7, or ACTB. The quantitation of integrated density of Fbxl7 and ACTB in (A–C) was shown in (D–F), respectively. Data are presented as mean ± standard error of Fbxl7/ACTB. *P < 0.05.

control mice (Fig. 2A and D). In 3-month-old rTg4510 mice, Fbxl7 protein level was significantly increased by 30% compared to control littermates (t-test, P = 0.024) (Fig. 2B and E). Ultimately, Fbxl7 expression was again found increased compared to control mice in 8-month-old rTg4510 mice (38%, P < 0.0001; Fig. 2C and F).

Publicly available human brain expression data sets

Fbxl7 was overexpressed (1.24-fold change; P = 3.0E-06) in 176 cases compared to 187 controls.[30] In the ADCs data set,[31] Fbxl7 was again overexpressed (1.06-fold change: P = 1.2E-05) in cases versus controls at a global level. Individuals where then stratified according to brain regions: Fbxl7 was significantly overexpressed in cases in entorhinal cortex and medial temporal gyrus (P = 0.002, 0.006 respectively). Full results for Fbxl7 expression data are presented in Table S3.

Discussion

The results of this study are twofold: the discovery of a novel locus not previously reported to be associated

with LOAD and the replication of six previously reported GWAS loci associated with LOAD in a unique population.

The finding of a genetic association between LOAD and *FBXL7* is important because it is a highly evolutionally conserved protein sharing a 98% sequence identity between mouse and humans. The biological role of Fbxl7 is not well understood, but F-box proteins constitute one of the subunits of E3 ubiquitin protein ligases involved in phosphorylation-dependent ubiquitination of proteins.[33] F-box proteins are also involved in several key biological functions including cell growth and differentiation, signal transduction, survival, and apoptosis[34,35] and interestingly in key AD-related pathological processes.[36,37] Their involvement in diverse conditions has been extensively studied in recent years (from cancer to neurological disorders such as early-onset Parkinson Disease[38]). The gene's overexpression in transfected cells displays proapoptotic effect in a dose and time-dependent manner.[34] Fbxl7 expression is thought to be regulated by another F-box protein, Fbxl18. Interestingly Fbxl18 has been identified as a direct target of *APP*, inhibiting neuronal differentiation.[39] *FBXL7* has been previously associated with several traits including the metabolic syndrome[40] and other vascular risk factors. Although never associated with LOAD at the genome-wide significant level, one SNP (rs11748700) within the *FBXL7* gene, was among the top hits in later meta-analysis showed a combined $P = 2.6E\text{-}06$.[41]

The association between LOAD and *FBXL7* was supported by differences in expression levels of the Fbxl7 protein in publicly available human brain microarray studies and in animal experiments. Transgenic mouse models of AD have been designed to act as surrogates of human pathological processes of neurodegeneration observed in AD (neurofibrillary tangles and senile plaques).[42] J20 mice, which show increased human Aβ fragments, develop several features typical of the human AD: synaptic loss, brain regions atrophy and cognitive impairment. rTg4510 mice have been designed to model the other pathological AD-signature, that is, neurofibrillary tangles and also show neuronal and synaptic loss. By providing evidence of altered expression of Fbxl7 in independent animal experiments that encompass both AD-signature neuropathological features, we speculate that the gene might act at different levels along the "amyloid cascade." The observed gene overexpression, confirmed in both human and animal data sets suggests a gain-of-function role in which there is altered protein-degradation activity resulting in accumulation of amyloid-β and tau proteins.

While we were unable to replicate the same *FBXL7* variant in the independent non-Hispanic whites and African-American data sets due to extreme differences in allele frequencies, we were able to confirm the association using different disease-associated SNPs in perfect LD ($D' = 1$) with rs75002042. Allelic heterogeneity across different ethnic groups in single-marker analyses is expected in complex diseases because of the occurrence of pathogenic mutations across multiple domains of disease genes (allelic heterogeneity) or because of the absence of these variants in some data sets or ethnic groups (locus heterogeneity). Variants found in GWASs likely tag the true functional variant; thus, the result for the tagging SNP may not replicate across studies even when the causal variant is effectively shared. These LD patterns depend on the ancestral background, as populations that have experienced bottlenecks (e.g., Caucasians) tend to show longer LD blocks and less evidence of recombination events when compared to Hispanics or African American.[43] Consequently, in admixed populations recombination events produce chromosomes that are mosaic of chromosomal regions originating from distinct ancestry and the so-called "flip-flop phenomenon" is well-known issue when interpreting trans-ethnic discordant results.[44] SNP rs75002042 in the *FBXL7* locus exhibits remarkable differences, with minor allele frequencies ranging from 1% in non-Hispanic Whites to 20% in African Americans.

Given that Caribbean Hispanics were not included in the IGAP[5] analysis we were inclined to replicate of at least some of the loci previously reported. We confirmed six known LOAD-susceptibility genes, either at a SNP or gene level. Interestingly, four of those (*FRMD4A*, *CELF1*, *FERMT2*, *SLC24A4-RIN3*) were only recently discovered using exceptionally large sample sizes. The SNP effects found in the GWAS results reported here, despite differences in the LD pattern, allele frequencies and ancestry, were in agreement with results reported in the IGAP study[5] in terms of direction and effect size. This further supports the role of those loci in the pathogenesis of LOAD, and validates the use of this cohort of individuals from another ethnic group.

This study does have limitations because of its unique ethnic composition, which makes it difficult to a similar independent cohort of similar ancestry to replicate the findings. While we cannot completely exclude the possibility that the results here are spurious because of the lack of an exact replication, we did provide evidence for *FBXL7* association using SNPs in complete LD in independent data sets in different ethnic groups further supported by expression studies in AD-transgenic mice lessening that conclusion. Additional confirmation (e.g., sequencing and functional analyses) will be mandatory to validate and fully understand the role of this novel candidate gene in the pathogenesis of LOAD. Taken together the results here suggest an alternative pathway in this complex disorder of aging.

Acknowledgments

This work was supported by federal grants from the National Institutes of Health and the National Institute on Aging, specifically P50 AG08702, R37 AG15473, RO1 AG037212; the Marilyn and Henry Taub Foundation and the Charles S Robertson Gift from the Banbury Fund (R. M.); R37 NS 29993 and the McKnight Brain Research Foundation (R. S. and M. E.); Canadian Institutes of Health Research (E. R., P. S. H.); Welcome Trust, Medical Research Council, Ontario Research Fund, and Alzheimer Society of Ontario (P. S. H.). We sincerely thank Lennart Mucke for providing J20 mice, Karen Ashe for providing FVB-Tg(tetO-Tau$_{P301L}$)4510 mice, Jada Lewis for providing 129S6-Tg(CaMKIIa-tTA) mice, Peter Davies for providing us CP27 antibodies, Mathieu Herman for collecting mouse brain tissues, and Helen Figueroa for maintaining mouse colonies.

Conflict of Interest

None declared.

References

1. Thies W, Bleiler L. 2013 Alzheimer's disease facts and figures. Alzheimers Dement 2013;9:208–245.
2. Slooter AJ, Cruts M, Kalmijn S, et al. Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study. Arch Neurol 1998;55:964–968.
3. Lambert J-C, Heath S, Even G, et al. Genome-wide association study indentifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 2009;41:1094–1099.
4. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 2009;41:1088–1093.
5. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013;45:1452–1458.
6. Tosto G, Reitz C. Genome-wide association studies in Alzheimer's disease: a review. Curr Neurol Neurosci Rep 2013;13:1–7.
7. Tang M-X, Cross P, Andrews H, et al. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. Neurology 2001;56:49–56.
8. Reitz C, Jun G, Naj A, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ε4, and the risk of late-onset Alzheimer disease in African Americans. JAMA 2013;309:1483–1492.
9. Vardarajan BN, Schaid DJ, Reitz C, et al. Inbreeding among Caribbean Hispanics from the Dominican Republic and its effects on risk of Alzheimer disease. Genet Med 2014;doi:10.1038/gim.2014.161.
10. Vardarajan BN, Faber KM, Bird TD, et al. Age-specific incidence rates for dementia and Alzheimer disease in NIA-LOAD/NCRAD and EFIGA families: National Institute on Aging Genetics Initiative for Late-Onset Alzheimer Disease/National Cell Repository for Alzheimer Disease (NIA-LOAD/NCRAD) and Estudio Familiar de Influencia Genetica en Alzheimer (EFIGA). JAMA Neurol 2014;71:315–323.
11. 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: 751–755.
12. Lee JH, Cheng R, Barral S, et al. Identification of novel loci for Alzheimer disease and replication of CLU, PICALM, and BIN1 in Caribbean Hispanic individuals. Arch Neurol 2011;68:320–328.
13. Sacco RL, Anand K, Lee H-S, et al. Homocysteine and the risk of ischemic stroke in a triethnic cohort the Northern Manhattan Study. Stroke 2004;35:2263–2269.
14. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:263–269.
15. Marchini J, Howie B, Myers S, et al. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 2007;39:906–913.
16. Consortium 1000 Genomes Project, others. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
17. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 2006;38:209–213.
18. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38: 904–909.
19. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2:e190.
20. Shriner D. Investigating population stratification and admixture using eigenanalysis of dense genotypes. Heredity 2011;107:413–420.
21. Zhu X, Li S, Cooper RS, Elston RC. A unified association analysis approach for family and unrelated samples correcting for stratification. Am J Hum Genet 2008;82:352–365.
22. Manichaikul A, Mychaleckyj JC, Rich SS, et al. Robust relationship inference in genome-wide association studies. Bioinformatics 2010;26:2867–2873.

23. Li M-X, Gui H-S, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. Am J Hum Genet 2011;88:283–293.
24. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19:1655–1664.
25. 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;43:436–441.
26. Ong RT-H, Teo Y-Y. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. Bioinformatics 2010;26:1269–1270.
27. Mucke L, Masliah E, Yu G-Q, et al. High-level neuronal expression of Abeta1-42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. J Neurosci 2000;20:4050–4058.
28. Santacruz K, Lewis J, Spires T, et al. Tau suppression in a neurodegenerative mouse model improves memory function. Science 2005;309:476–481.
29. Mayford M, Bach ME, Huang Y-Y, et al. Control of memory formation through regulated expression of a CaMKII transgene. Science 1996;274:1678–1683.
30. Myers AJ, Gibbs JR, Webster JA, et al. A survey of genetic human cortical gene expression. Nat Genet 2007;39: 1494–1499.
31. Liang WS, Reiman EM, Valla J, et al. Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. Proc Natl Acad Sci USA 2008;105:4441–4446.
32. Lambert J-C, Grenier-Boley B, Harold D, et al. Genome-wide haplotype association study identifies the FRMD4A gene as a risk locus for Alzheimer's disease. Mol Psychiatry 2012;18:461–470.
33. Rodrigues-Campos M, Thompson BJ. The ubiquitin ligase FbxL7 regulates the Dachsous-Fat-Dachs system in Drosophila. Development 2014;141:4098–4103.
34. Liu Y, Lear T, Zhao Y, et al. F-box protein Fbxl18 mediates polyubiquitylation and proteasomal degradation of the pro-apoptotic SCF subunit Fbxl7. Cell Death Dis 2015;6:e1630.
35. Bosch JA, Sumabat TM, Hafezi Y, et al. The Drosophila F-box protein Fbxl7 binds to the protocadherin Fat and regulates Dachs localization and Hippo signaling. Elife 2014;3:e03383.
36. Kim M-Y, Mo J-S, Ann E-J, et al. Regulation of Notch1 signaling by the APP intracellular domain facilitates degradation of the Notch1 intracellular domain and RBP-Jk. J Cell Sci 2011;124:1831–1843.
37. Watanabe T, von der Kammer H, Wang X, et al. Neuronal expression of F-box and leucine-rich-repeat protein 2 decreases over Braak stages in the brains of Alzheimer's disease patients. Neurodegener Dis 2012;11:1–12.
38. Di Fonzo A, Dekker M, Montagna P, et al. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. Neurology 2009;72:240–245.
39. Shu R, Wong W, Ma Q, et al. APP intracellular domain acts as a transcriptional regulator of miR-663 suppressing neuronal differentiation. Cell Death Dis 2015;6:e1651.
40. Zhang Y, Kent JW, Olivier M, et al. A comprehensive analysis of adiponectin QTLs using SNP association, SNP cis-effects on peripheral blood gene expression and gene expression correlation identified novel metabolic syndrome (MetS) genes with potential role in carcinogenesis and systemic inflammation. BMC Med Genomics 2013;6:14.
41. Shi H, Medway C, Bullock J, et al. Analysis of Genome-Wide Association Study (GWAS) data looking for replicating signals in Alzheimer's disease (AD). Int J Mol Epidemiol Genet 2010;1:53.
42. Bryan KJ, Lee H, Perry G, et al. Transgenic mouse models of Alzheimer's disease: behavioral testing and considerations. 2009.
43. Rosenberg NA, Huang L, Jewett EM, et al. Genome-wide association studies in diverse populations. Nat Rev Genet 2010;11:356–366.
44. Lin P-I, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: the flip-flop phenomenon. Am J Hum Genet 2007;80:531–538.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Supplemental file.

Table S1. Genotype characteristics for the EFIGA, WHICAP and NOMAS studies. Individuals derived from the EFIGA and WHICAP studies were genotyped over the past few years in three consecutive batches. On the contrary, all NOMAS individuals were genotyped at one time. Each batch was imputed singularly although we used the same algorithm that is, the March 2012 reference panel from 1000 Genomes (1000G) – build 37 and the IMPUTE2 software (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html [Marchini et al. 2007]). The resulting imputed datasets were then merged using the GTOOL software (http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html); finally, imputed SNPs were included in the combined dataset only if present in all four cohorts.

Table S2. SNPs allele frequencies for genome-wide significant variants along with IGAP meta-analysis variants that were replicated in our GWAS; allele frequencies are compared to those of 1000G reference superpopulations (HISP = Caribbean Hispanics; EUR = European; AFR = African; AM-ADMX = American Admixed; EAS = East Asian).

Table S3. *FBXL7* and *CACNA2D3* expression in LOAD brains compared to healthy brains. To determine if expression levels differ between LOAD and unaffected brains, we performed analysis of variance (ANOVA). Kruskal–Wallis test and Monte Carlo estimates were used for the regional comparisons in the Liang et al. sample. Prior to analysis, $\log_{10}$-transformed rank invariant normalized the expression data.

Figure S1. Quantile-quantile plot with the exclusion of the APOE locus (chr19: 44,000,000–47,000,000) and the genome-wide significant loci.

Figure S2. Manhattan plot for the case-control GWAS. Each dot represents one SNP and its projection on the *Y* axis indicates the associated *P*-value. The red line identifies SNPs that passed the threshold for genome-wide significance ($P \leq$ 5E-08) to account for multiple testing correction.

Figure S3. Regional plots generated with the Locuszoom software (http://csg.sph.umich.edu/locuszoom/) for the *FBXL7* locus employing the African reference panel (FBXL7 – AFR) and the European reference panel (FBXL7 – EUR) from the 1000G project to estimate the LD pattern. SNP in purple is the top associated variant.

Figure S4. Ancestry estimation in the unrelated case-control sample using the supervised admixture algorithm as implemented in the ADMIXTURE software (Red color indicates European ancestry, green the African ancestry and blue color the Native American ancestry).

