Award Number: W81XWH-12-1-0013

TITLE: Whole Exome Analysis of Early Onset Alzheimer's Disease

PRINCIPAL INVESTIGATOR: Margaret Pericak-Vance

CONTRACTING ORGANIZATION: University of Miami, 1400 NW 10th Avenue, Miami, FL 33136

REPORT DATE: OCT 2018

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCT 2018	FINAL	1 March 2012 – 30 July 2018

4. TITLE AND SUBTITLE:

Whole Exome Analysis of Early Onset Alzheimer's Disease

5a. CONTRACT NUMBER
W81XWH-12-1-0013

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Margaret A. Pericak-Vance, Ph.D.

E-Mail: mpericak@med.miami.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Miami, 1400 NW 10th Avenue, Miami, FL 33136

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The primary focus toward identification of Alzheimer disease (AD) risk genes over the past five years has been testing the common disease common variant (CDCV) hypothesis through the use of genome-wide association studies (GWAS) in late onset Alzheimer disease (LOAD). While common variation clearly plays a role in AD there is a growing realization that the CDCV hypothesis is unlikely to explain all the genetic effect underlying AD. One alternative hypothesis invokes multiple rare variants (RV) in one or more genes, each with stronger individual effects than CDCV genes. We designed this project to test the rare variant hypothesis in AD by examining those cases with the most severe phenotype as determine by early onset (EOAD, cases with AAO < 60 years). Although there are three known EOAD genes (PS1, PS2 and APP) they account for only ~60-70% of familial EOAD and even less of sporadic EOAD. Thus, the majority of the genetics of EOAD remains unknown. Until now, large extended families with AD in multiple generations were necessary to identify variants of significant effect contributing to AD risk, however, with the advent of new genomic technologies such as high-throughput sequencing technology, small family aggregates and isolated cases, particularly those with an extreme phenotype of the disorder (such as early onset) can be used. Thus, we will utilize whole exome high-throughput sequencing to identify high risk AD variants that we will further characterize with respect to AD. We will examine both Caucasian and Caribbean Hispanic AD populations. Our two pronged approach includes structural characterization at the DNA level (Dr. Pericak-Vance), and analysis of Caribbean Hispanics (Dr. Richard Mayeux). Comparing across populations will be extremely useful. Specifically, high priority RVs identified through the whole exome analysis will be further explored with multiple strategies. We will also genotype the interesting variants in a large sample of late-onset (LOAD) cases to examine their involvement in all AD. We will thus prepare a list of high priority candidates for additional follow-up and functional analysis.

.

15. SUBJECT TERMS

NONE LISTED

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	23	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

INTRODUCTION:

The primary focus in the identification of Alzheimer disease (AD) risk genes has focused on the common disease common variant (CDCV) hypothesis using genome-wide association studies (GWAS) in late onset Alzheimer disease (LOAD). It is clear that common variants play an important role in AD, the CDCV hypothesis can't fully explain the genetic factors underlying AD. As an alternative, recent genetic studies have focused on the identification of multiple rare variants (RV) in one or more genes, each with string effect sizes. To that end, the current study was designed to test the rare variant hypothesis in AD by examining those cases with the most severe phenotype as determine by early onset (EOAD, cases with AAO < 60 years). There are three known EOAD genes – *Presenilin 1* (*PS1*), *Presenilin 2* (*PS2*), and *Amyloid precursor protein* (*APP*) – that account for ~60- 70% of familial EOAD cases and fewer in sporadic EOAD and, as such, the majority of EOAD genes remain to be identified. To that end, we will utilize whole exome next generation sequencing (NGS) to identify high risk AD genetics variants. We will examine both Caucasian and Caribbean Hispanic AD populations. Our two pronged approach includes structural characterization at the DNA level (Dr. Pericak-Vance), and analysis of Caribbean Hispanics (Dr. Richard Mayeux). Comparing across populations will be extremely useful. Specifically, high priority RVs identified through the whole exome analysis will be further analysis, including bioinformatics and computational analysis, genotyping of variants in a large sample of late-onset (LOAD), as well as, functional characterization using patient-specific induced pluripotent stem cells (iPSCs). Patient specific iPSC derived from EOAD patient samples bearing genetic variants of interest will be developed and differentiated into forebrain neurons that will be characterized for markers of AD pathogenesis, including expression of pathogenic amyloid beta and tau isoforms.

BODY:

ACCOMPLISHMENTS:

Phase I – Sequencing and Validation of Variants of Interest

WES and variant prioritization

Whole exome sequencing (WES), quality control and variant calling, variant annotation, and variant filtering is complete on 55 samples submitted by Columbia University to the University of Miami. Additionally, WES and analysis of 51 samples from 46 multiplex families from The University of Miami and Vanderbilt University is complete. Identity-by-descent analysis of Hispanic families was also performed. Following these analyses, comparison of the candidate variants/genes shared across Hispanic families and NH-white cases was done. From these analyses, a list of 125 unique variants was prioritized for follow-up genotyping.

A brief overview of how each family was filtered individually and how variants for typing were prioritized follows:

1) Quality Filter per individual WES sample: VQSLOD>0, PL Score>100, Read Depth>6
2) Annotation of remaining variants with ANNOVAR
3) Remove variants with MAF>0.001 in EVS_6500si and 1000G2012mar_all and MAF>0.01 in HIHG internal controls
4) Keep variants with Autosomal dominant and X-linked dominant segregation in family
4) Exclude variant if not missense, Splicing, Stopgain, Stoploss, Nonframeshift Indel, or Frameshift Indel in refSeq gene annotation, Ensemble gene annotation, or UCSC Known gene annotation
5) Filter on deleteriousness based on a) damaging score in any of these 7 programs programs: Sift, Polyphen2_HDIV, LRT, MutationTaster, MutationAssessor, or FATHMM and b) conservation based on a conserved score in any of these 3 programs: GERP, SiPhy or PhyloP
7) Apply IBD sharing results and require 100% sharing in Hispanic families with enough GWASed individuals
8) Genotype any variant passing above filters and in a known EOAD or LOAD
9) Interrogate shared variants and variants in shared genes across Hispanic Families and between Hispanic and NH-White Families by screening them for existence and potentially too high a MAF in dbSNP, EVS, 1000G updates, specific 1000G populations (EA, AA, AMR and ASN, and any population in UCSC), and cg69 (69 complete genomics exomes). Because of the large amount of candidate genes generated from filtering of the NH-White cases, a variant from the comparison of Hispanic and NH-White candidates was only carried forward for genotyping if the variant/gene passed this screening and was in 2+ Hispanic families and 2+ NH-White cases. Additionally, variants/genes still in 2+ Hispanic families after the screening were carried forward for genotyping.
8) Additional variants were selected by applying a 'secondary filter' to the Hispanic families in order to reduce single variant per family candidates:
---remove any SNV with an rs# in dbSNP129-dbSNP137
---remove all indels
---remove families with greater than 50 variants remaining (families 1,171,386 and 419)
---keep only variants predicted to be damaging in 3 or more of the 7 prediction programs used
---NOTE: Candidate variants for the four removed families were selected based on shared variants/genes with other families.

Follow-up Genotyping of Top Candidate Variants

261 Hispanic familial subjects from 19 pedigrees (145 affecteds and 116 unaffecteds) and 500 Hispanic non-familial subjects (382 healthy controls and 118 sporadic EOAD cases) were genotyped for these 125 top candidate variants. 101 of the variants passed all QC filters (13 variants failed genotyping and 11 were monomorphic in the dataset). For analysis of results of this follow-up genotyping we: 1) estimated familial and population frequencies of the variants in our follow-up cohort and 2) tested single SNV association with AD with 2 models using generalized estimation equations (GEE):

M1) AD~SNV+AGE+SEX
M2) AD~SNV+AGE+SEX+APOE

20 top candidate variants were identified from this follow-up genotyping, including a 44 base-pair deletion in

ABCA7 that was further genotyped and Sanger sequenced separately. The other 19 variants include 8 SNVs that show perfect segregation with AD status in the families and are absent in population controls (Table 1). These variants are in the genes *MYO3A, AAAS, DICER1, YIPF1, ACAP1, LLGL2, BPIFB2*, and *ABCG2*. An additional 11 variants were identified as follow-up candidates based on them showing near complete segregation (absent in one or a few familial cases) and being absent in all familial and sporadic controls. These variants are in the genes *GPR26*, *ERCC*6, *OR5M9, DNAH3, MYOCD, KIF17, TICRR, PLXNB2, LAMA2, SNRNP48, and GLB1L2.* These top 19 variants were then genotyped in large cohorts of Hispanics (1621 cases and 884 controls) (Table 2), African-Americans (157 familial cases, 400 sporadic cases, and 942 unrelated cases) and Non-Hispanic Caucasians (2,377 familial cases, 739 sporadic cases, and 600 unrelated cases). Genotyping in AA identified 1 case with the *DNAH3* variant; 3 controls with the *TICRR* variant (Age-of- Exams of 83, 62, and 67), 1 control with the *SNRNP48* variant (Age-of-Exam of 70), and 1 control with the *LAMA2* SNV (Age-of-Exam of 70). Genotyping in NHW cases and controls identified 1 familial case with the KIF17 variant (Age-of-onset of 75) and another case in *PLXNB2* (Age-of-onset of 51). NOTE: Assays for *MYOCD, ACAP1, LLGL2 DICER1*, and *GPR26* could not be designed for follow-up genotyping in AA and NHW. Options for genotyping for these variants are under consideration.

Table 1. 19 candidate variants from stage 1 validation genotyping of 125 candidate variants.

CHR	POS	GENE	+ Family Cases	+ Family Ctrls	+ Sporadic Cases	+ Sporadic Controls	- Family Cases
10	26243836	*MYO3A*	4	0	0	0	0
12	53714383	*AAAS*	5	0	0	0	0
14	95574334	*DICER1*	5	0	0	0	0
1	54354580	*YIPF1*	5	0	0	0	0
17	7249740	*ACAP1*	5	0	0	0	0
17	73552150	*LLGL2*	7	0	0	0	0
20	31606097	*BPIFB2*	5	0	0	0	0
4	89042878	*ABCG2*	5	0	0	0	0
10	1.25E+08	*GPR26*	3	0	0	0	1
10	50667132	*ERCC6*	4	0	0	0	1
11	56230523	*OR5M9*	6	0	1	0	1
16	21071613	*DNAH3*	2	0	1	0	1
17	12626229	*MYOCD*	5	0	0	0	1
1	21031194	*KIF17*	5	0	0	0	2
15	90145195	*TICRR*	5	0	0	0	2
22	50727490	*PLXNB2*	5	0	0	0	2
6	1.3E+08	*LAMA2*	5	0	2	0	2
6	7605638	*SNRNP48*	5	0	0	0	2
11	1.34E+08	*GLB1L2*	4	0	0	0	3

Table 2. Results of follow-up genotyping of top 19 Hispanic EOAD variants in 1621 Hispanic Cases and 884 Hispanic Controls

Chr:Position	Gene	1621 Hispanic Cases MAF	884 Hispanic Controls MAF
10:26243836	*MYO3A*	0.0009482	0
12:53714383	*AAAS*	0	0
14:95574334	*DICER1*	0	0
1:54354580	*YIPF1*	0	0
17:7249740	*ACAP1*	0.0006324	0.001705
17:73552150	*LLGL2*	0.003519	0.005143
20:31606097	*BPIFB2*	0	0
4:89042878	*ABCG2*	0	0
10:125426036	*GPR26*	0	0
10:50667132	*ERCC6*	0.001261	0.002286
11:56230523	*OR5M9*	0.002527	0.001708
16:21071613	*DNAH3*	0	0
17:12626229	*MYOCD*	0	0

1:21031194	*KIF17*	0.0006333	0.001139
15:90145195	*TICRR*	0.0006321	0.001139
22:50727490	*PLXNB2*	0	0
6:129714360	*LAMA2*	0.001262	0.0005675
6:7605638	*SNRNP48*	0	0
11:134241025	*GLB1L2*	0	0

Follow-up genotyping of the 44 base pair deletion in *ABCA7* (rs142076058) found segregating in Hispanic Family 380 and the Reitz et al. 2013 (1) *ABCA7* risk SNP (rs115550680) was completed in cohorts of NHW and AA AD cases and controls. Results show the deletion is very rare in Non-Hispanic White cases and controls (0.12%). Testing in AA cases and controls, adjustmenting for age, sex, and APOE status, found the deletion to be significantly associated with disease (p=0.0002, OR=2.13 [95% CI:1.42-3.20]). The association was replicated in an independent dataset (p=0.0117, OR=1.65 [95% CI:1.12-2.44]). Joint analysis resulted in an effect size (OR) estimate = 1.81 ([95% CI:1.38-2.37] p=1.414x10-5). The deletion is common in both AA cases (15.2%) and AA controls (9.74%). Linkage disequilibrium between the deletion and the Reitz et al. 2013 risk SNP is high at r2 = 0.921 and D' = 0.975.

Non-Hispanic White Cases Only Analyses

To search for rare variants contributing to risk for EOAD we performed Whole-Exome Sequencing (WES) in 50 Caucasian EOAD cases screened negative for *APP, PSEN1, and PSEN2*. Variant filtering for rare (MAF<0.1% in ExAC database) functional (nonsynonymous or loss-of-function(LOF)), damaging variants was performed. Damage prediction was performed using a Combined Annotation Dependent Depletion (CADD) score (2), with scores ≥ 10 considered damaging. Rare, damaging variants shared by multiple cases (+2) were then selected for follow-up protein-protein interaction analysis with known or suspected EOAD genes (*APP, PSEN1, PSEN2, GRN, MAPT, TREM2, SORL1*) using the program STRINGdb (3). This analysis identified 5 genes with the same rare, potentially damaging nonsynonymous or LOF variant in two or more EOAD cases and evidence for protein-protein interaction with a known EOAD gene (Table 3 and Figure 1). Several other cases have a rare nonsynonymous or LOF, potentially damaging variant in another variant in these 5 genes.

The 5 genes implicated are: *HSPG2* (interacts with *GRN* and *APP*), *CLSTN1* (interacts with *PSEN1* and *APP*), *DOCK3* (interacts with *PSEN1* and *PSEN2*), *PARK2* (interacts with *MAPT, PSEN1*, and *APP*), and *OGT* (interacts with MAPT) (Figure 1). Six cases have a variant in *HSPG2*, a gene in a LOAD susceptibility region and potentially involved in amyloidogenesis and tau aggregation in AD (4-6). Three cases have a variant in *DOCK3*, a gene shown to regulate amyloid-β secretion, and associated with neurofibrillary tangles in AD brains (7,8). Two cases have shared variants in *CLSTN1*. Disruption of calsyntenin-1-associated axonal transport of APP by mutations in *CLSTN1*, a known APP interactor (9,10), have been identified as a potential pathogenic mechanism of Alzheimer's (11). Moreover, *CLSTN1*'s potential as a regulator of synapse formation and neuronal development suggests other mechanisms through which it could be involved in development of dementia (12). Interestingly, *CLSTN1* interacts with another newly identified candidate gene from this analysis, *OGT,* which was found to have 12 EOAD cases carrying two separate frameshift insertions at the same position on Chromosome X (X: 70767666). Numerous studies exist linking *OGT* to neurodegeneration, including a study supporting its therapeutic potential due to its ability to prevent protein aggregation including reduction of formation of tau oligomers [13], and a study showing increased biochemical levels of *OGT* lead to slower cognitive decline and amyloid plaque formation in mice [14]. Finally, though variants in *PARK2* are the most frequent cause of autosomal recessive early-onset Parkinson's disease and juvenile Parkinson disease, Parkin has been shown to promote intracellular Abeta1–42 clearance [15], is upreregulated in AD brains, and colocalizes with classic senile plaques and amyloid-laden vessels in AD brains [16], hinting at its potential involvement in Alzheimer's as well. Validation of these results using Sanger sequencing is underway.

Table 3. Rare (MAF<0.001), nonsynonymous or loss-of-function variants (LOF) found in two or more EOAD cases. Additional rare, missense or LOF variants in these genes are also listed.

Chr:Position	Ref/Alt Allele	rsID	Gene	Consequence	N Affected	MAF (ExAC)	CADD	STRINGdB Interaction
1:9795622	C/T	.	CLSTN1	missense	2	8.95E-05	15.92	APP, PSEN1, OGT
1:22211078	T/G		HSPG2	missense	2	N	16.49	APP, GRN
1:22163463	G/A	rs200225298	HSPG2	missense	1	2.47E-04	11.89	APP, GRN
1:22181841	C/T	.	HSPG2	missense	1	5.70E-05	22.9	APP, GRN
1:22183791	G/A	.	HSPG2	missense	1	8.13E-05	26.4	APP, GRN
1:22205523	C/T	rs150650673	HSPG2	missense	1	1.14E-04	28.6	APP, GRN
3:51251601	G/A	rs199600118	DOCK3	missense	2	5.80E-04	29.5	PSEN1, PSEN2
3:51312575	G/C	.	DOCK3	missense	1	N	29	PSEN1, PSEN2
6:161969982	G/+C	.	PARK2	frameshift	1	8.13E-06	35	APP, MAPT, PSEN1
6:161990390	C/G	rs72480423	PARK2	missense	3	1.95E-04	22.4	APP, MAPT, PSEN1
X:70767666	T/+C	.	OGT	frameshift	8	1.72E-04	14.86	MAPT, CLSTN1
X:70767666	T/+CC	.	OGT	frameshift	4	1.72E-04	14.86	MAPT, CLSTN1

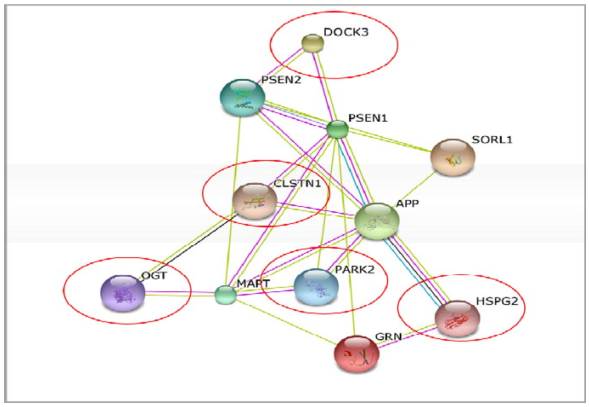


Figure 2. STRINGdb network analysis of top Non-white Hispanic EOAD candidate genes using known or suspected EOAD genes(APP, PSEN1, PSEN2, TREM2, MAPT, TREM2, SORL1, and GRN) as seed nodes. 'Strings' between genes represent evidence of protein-protein interaction between linked genes.

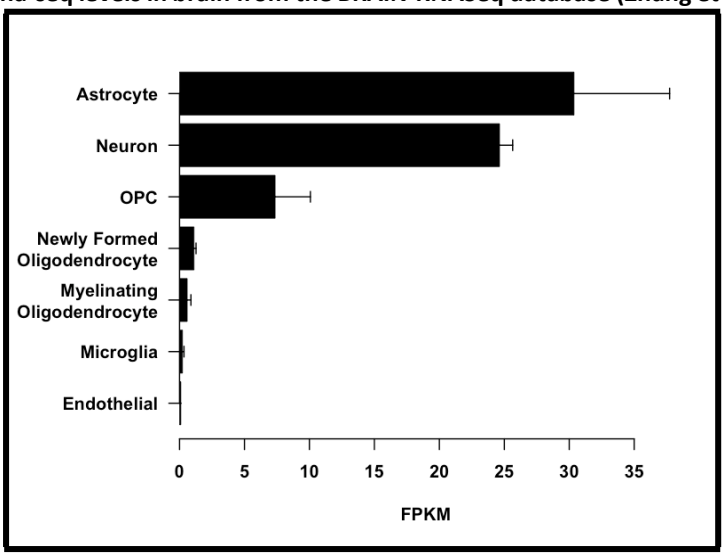
We additionally found several rare coding variants in known or suspected EOAD genes (Table 3), and are investigating a potential link to Parkinsonism and SORL1 (see Cuccaro ASHG 2015 abstract below).

Table 3. Variants in non-Hispanic White Cases in known EOAD genes.

N Cases	Age	Chr	Start	Ref	Alt	Gene
2	61,61	17	44101427	C	T	MAPT
1	52	17	44101427	C	T	MAPT
1	54	14	73637653	C	T	PSEN1
1	56	14	73637653	C	T	PSEN1
1	50	14	73664774	C	G	PSEN1
1	59	11	121414334	C	T	SORL1
1	55	11	121498300	C	T	SORL1
1	48	1	227071475	C	T	PSEN2
1	48	1	227075813	A	G	PSEN2

We also completed analysis of a comparison between the Alzheimer's Disease Genetics Consortium (ADGC) early onset Alzheimer's disease exome chip case-control association study and the WES produced from this project. We first updated the association analysis to include a 5th cohort, bringing the total N of the sample to 1,292 cases and 5,625 controls. Analysis comparing the rare, high consequence (missense, non-frameshift, loss-of-function) variants in the NHW WES dataset to the ADGC exome chip association results was then conducted. Briefly, nine genes are genome-wide significant at a Bonferonni correction for 7,249 genes tested ($P=6.90 \times 10^{-6}$), including *PSD2* ($P=6.98 \times 10^{-7}$), an endocytic gene with 2 rare, missense variants present in two separate NHW EOAD cases (Table 3). Preliminary bioinformatics analyses shows both *PSD2* variants in the WES cases to have high CADD scores of 27.4 and 28.5 (above 15 considered damaging) (Kirchner et al. 2014). Additionally, the gene, which is exclusively expressed in brain according to The Human Protein Atlas, is significantly overexpressed in both neurons and astrocytes according to the database Brain-RNASeq (Figure 2) (Zhang et al. 2014). Two genes with rare, segregating variants in the Hispanic families (PER3 and PCDHB11) were found to be genome-wide significant as well (1.74×10^{-7} and 8.92×10^{-7}). Additionally, the gene *IL16*, in which WES found 2 NHW cases and 2 Hispanic families with rare, missense variants, had suggestive significance in the exome chip study ($P=8.33 \times 10^{-4}$). Results have been incorporated into a manuscript.

***PSD2* rna-seq levels in brain from the BRAIN-RNASeq database (Zhang et al. 2014).**



Top Results, ranked by P-value, of ADGC Exome Chip Analysis. Genes above the red line are genome-wide significant at a Bonferonni correction for 7249 genes tested (P=6.90 x 10-3); The blue line represents significance for 2+ non-Hispanic White (NHW) EOAD cases with a rare, damaging variant in the same gene (910 genes tested; P=5.49 x 10-5); The yellow line represent significance for 2+ Hispanic families with a segregating rare, damaging variant in the gene same gene (73 genes tested; 6.85 x 10-4).

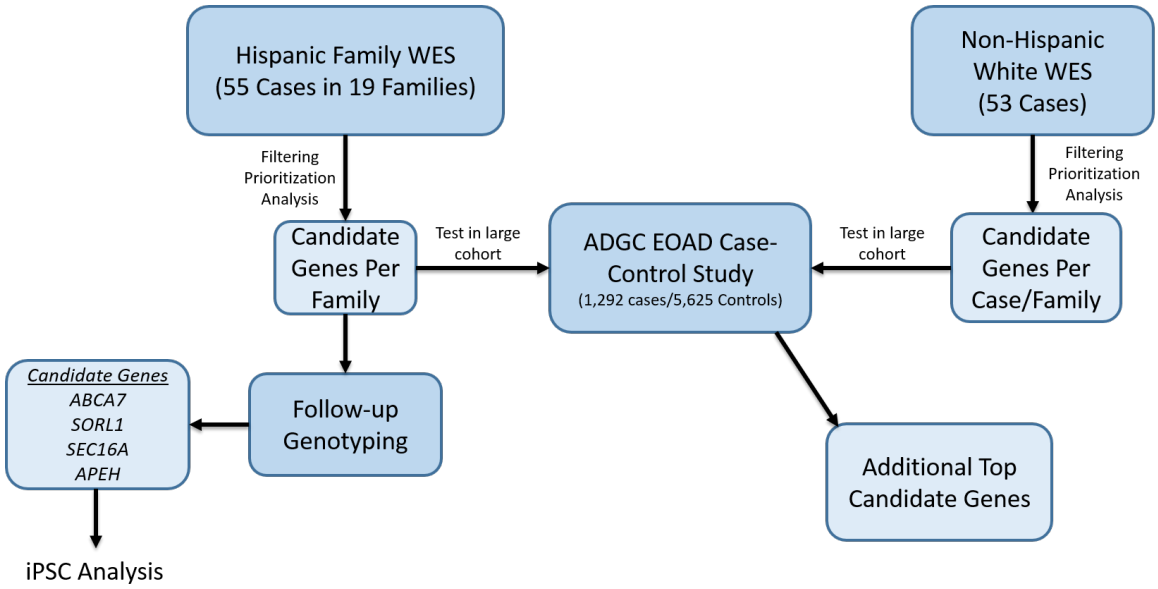
Gene	P-Value	N Rare SNPs Tested	N NHW Cases*	N Hispanic Families**
RFTN1	7.35E-08	9		
PER3	1.74E-07	29		1
SH2B3	2.08E-07	13		
ZFYVE9	3.52E-07	25		
MYEOV2	4.86E-07	5		
C1GALT1	5.02E-07	3		
PSD2	6.98E-07	17	2	
PCDHB11	8.92E-07	10		1
BSG	1.02E-06	6		
PADI1	2.57E-05	16		
MUC17	3.70E-05	66		

TGFB1	5.31E-05	3			
PKD1	1.52E-04	18			
LONP1	1.61E-04	23			
NPC1L1	2.01E-04	31			
RBFOX1	2.38E-04	10			1
ABR	2.43E-04	6			
EXD3	4.61E-04	49			
KLHDC7B	4.73E-04	4			
P2RY4	5.47E-04	11	2		
MEGF8	6.04E-04	24	3		
EMID1	6.07E-04	5			
ADAM17	7.85E-04	15			
FBF1	8.18E-04	18	4		
PBLD	8.18E-04	7			
C20orf123	8.22E-04	5			
CEACAM20	8.29E-04	16			
IL16	8.33E-04	20	2		2
MAPK11	9.87E-04	2			

*Number of NHW cases with a rare, damaging variant in the gene

**Number of Hispanic families with a rare, damaging variant in the gene

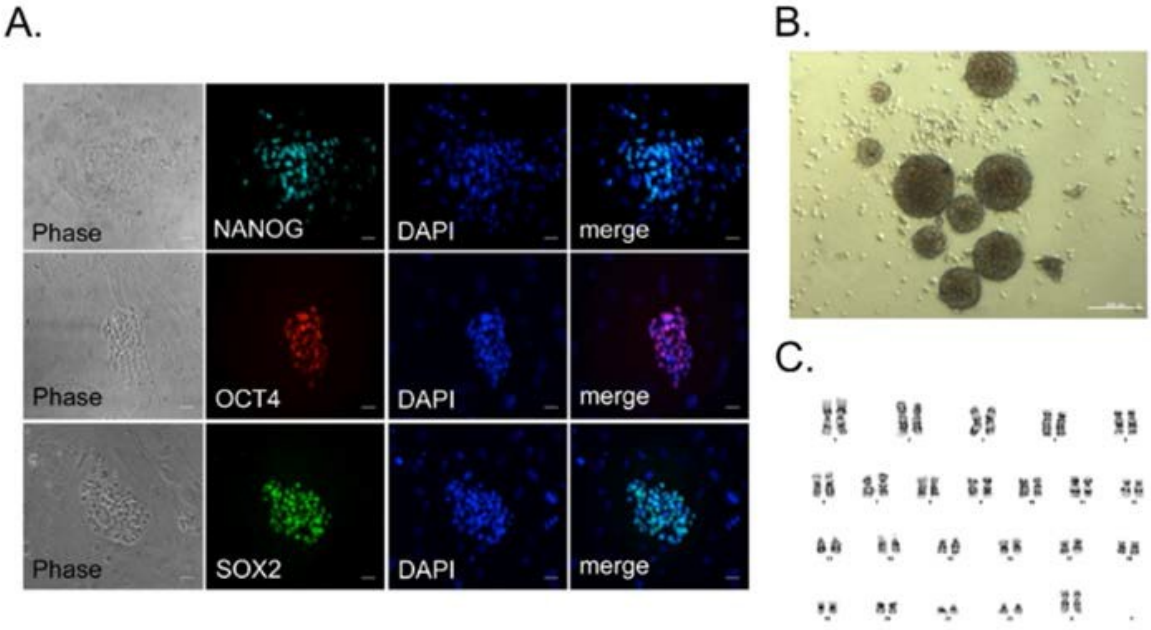
We have completed analysis of a comparison between the Alzheimer's Disease Genetics Consortium (ADGC) early onset Alzheimer's disease exome chip case-control association study and the WES produced from this project. Nine genes are genome-wide significant at a Bonferonni correction for 7,249 genes tested, including *PSD2* (P=6.98 x 10^{-7}), an endocytic gene with 2 rare, missense variants present in two separate NHW EOAD cases. Below is the flow chart of the study, with the most interesting genes and variants being funneled into the iPSC analysis.



Phase II – AD iPSC Functional Studies

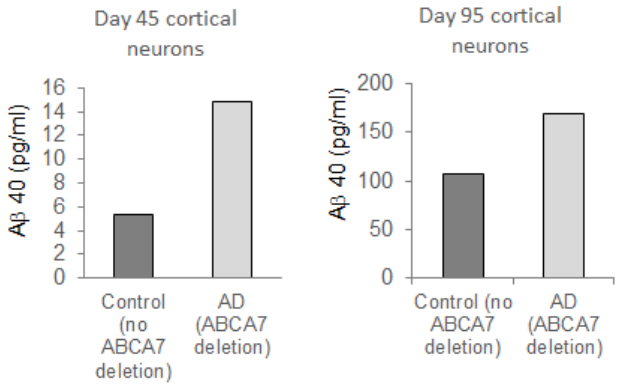
During the current funding period, we were able to collect peripheral blood mononuclear cells (PBMCs) from the whole blood of AD individuals, as well as, race and gender-matched control individuals. These PBMCs were reprogrammed in iPSC through the transient overexpression of the Yamanaka factors – OCT4, SOX2, KLF2, and c-MYC using the Sendai virus system. We have derived multiple lines from non-hispanic white individuals bearing variants in the SORL1 or TTC3 gene, African American individuals bearing an ethnic-specific deletion in the ABCA7 gene, and Caribbean-Hispanic (Dominican Republic) individuals bearing variants in the SEC16A gene. These lines have been characterized for their

pluripotency by immunocytochemistry (ICC), functional pluripotency through embryoid body formation, and karyotype analysis to ensure the stability of the genomes.



Validation of pluripotency of AD iPSC bearing mutations in the ABCA7 gene. A. Immunocytochemistry for pluripotency factors (Nanog, Oct4 and Sox2). B) These iPSC have the capacity to form embryoid bodies (a test of pluripotency) and were found to have a normal karyotype (C).

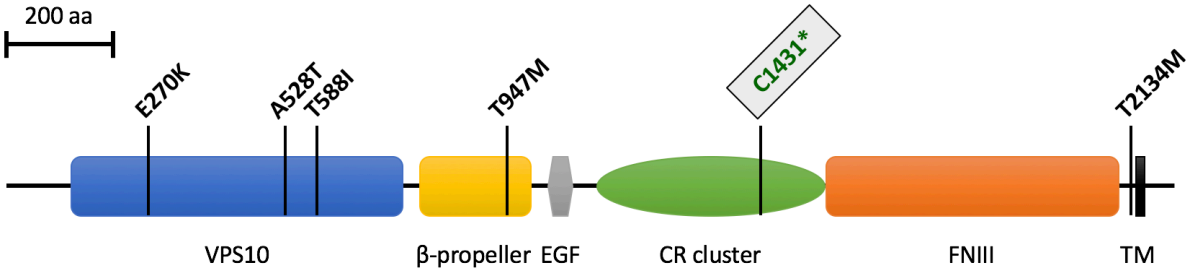
We have begun the differentiation of the iPSC lines into forebrain neurons using a multistep approach beginning with the formation of neurospheres, the plating of the neurospheres on poly-L-Ornithine/laminin to form neural rosettes, the formation and expansion of neural progenitor cells. The neural progenitor cells will then be replated on poly-D-lysine/laminin and differentiated into forebrain neurons (as determined by staining for the expression of the appropriate markers). These neurons were analyzed at different time post initiation of differentiation to identify the optimal timing for the analysis of the different amyloid beta species (Aβ40 and Aβ42) from the culture supernatant and tau and phospho-tau species from intracellular lysates of the iPSC derived neurons. In our preliminary results, we found that there were elevated levels of Aβ40 in the ABCA7 deletion bearing sample.



Amyloid beta 40 levels in iPSC-derived neurons at days 45 and 90 post initiation of differentiation in an AD-specific neurons (ABCA7 deletion bearing) compared to matched controls.

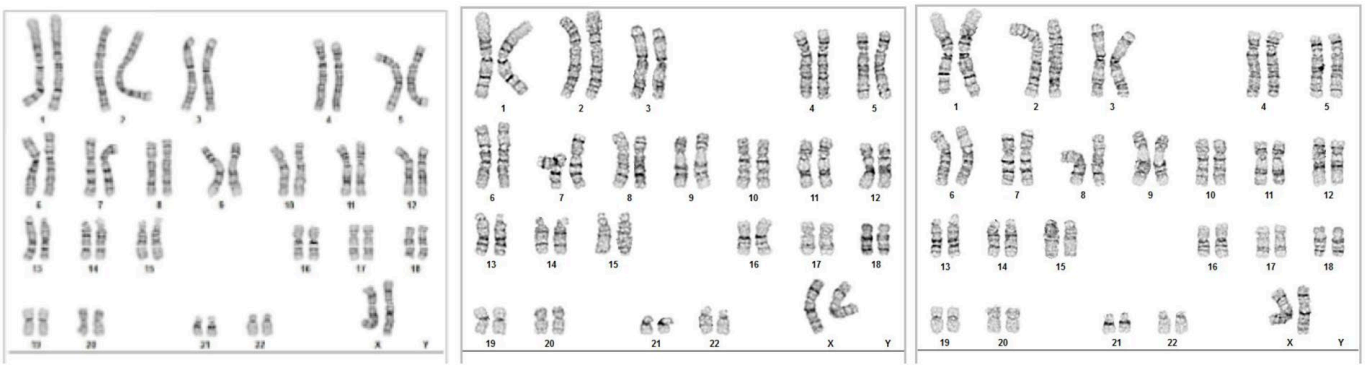
For the generation of additional iPSC lines, we have collected PBMCs from EOAD non-Hispanic white individuals bearing an alteration in the *SORL1* (shown below), African American individuals carrying an ethnic-specific deletion in the *ABCA7* gene, and Caribbean-Hispanic (Dominican Republic) individuals bearing variants in the *SEC16A* gene. Two iPSC lines has been made from individuals with the *SORL1* from two siblings, one with AD and another with mild cognitive impairment, often a precursor to AD. In addition, iPSC lines have ben generated from an AD patient another with the *SEC16A* alteration, and two patients carrying with the *ABCA7*

deletion.

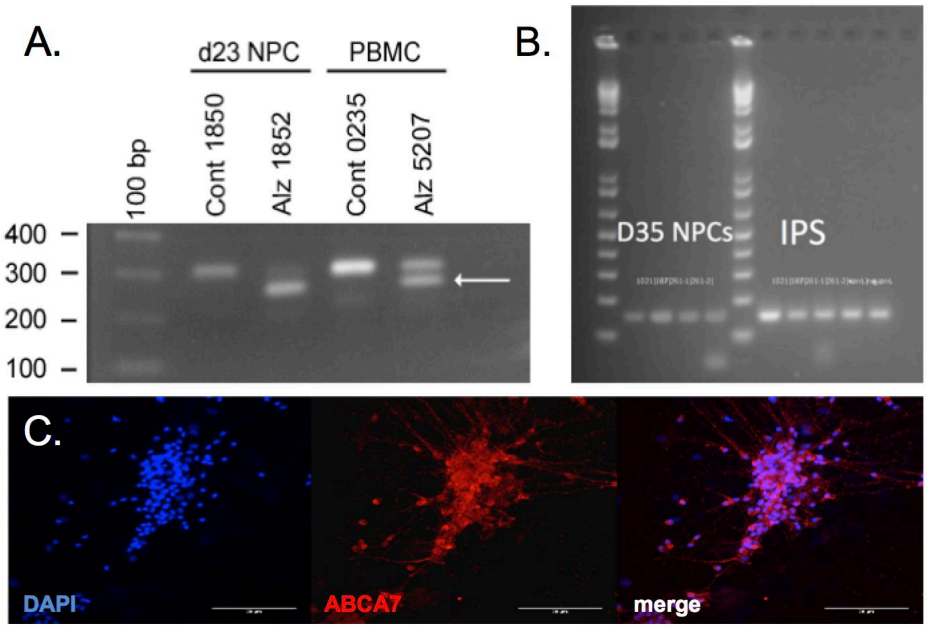


The frameshift alteration in the SORL1 protein falls in the CR-cluster, which is essential for interaction with APP. Removing the CR-cluster abolishes the protection against APP processing (Mehmedbasic, et al, 2015).

Two clones were isolated from each of these samples and tested via karyotyping to ensure that no chromosomal abnormalities arose during the reprogramming process (see below). Additional clones and a pool of clones were also cryopreserved for additional potential clones of each line for the future.

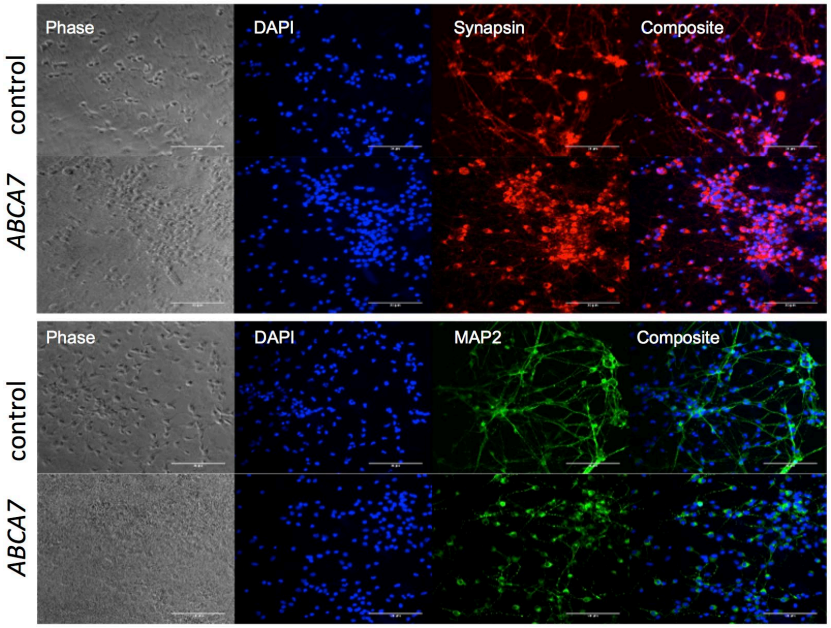


Karyotype analysis from iPSC clones of derived from a female with a deletion in the *ABCA7* (left), the individual with the variant in *SORL1* (middle), and an individual that carries an alteration in the *SEC16A* gene (right). All karyotypes are normal.



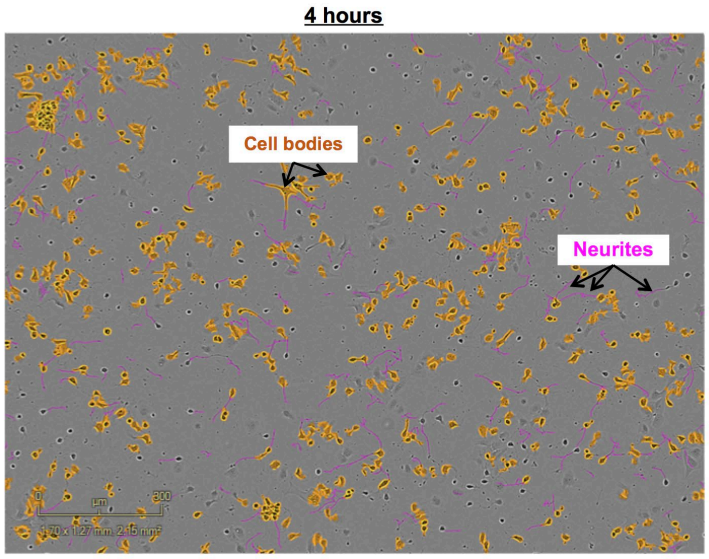
We were able to determine that the *ABCA7* and *SORL1* lines both expressed these genes during neuronal differentiation. We isolated RNA from these patient-derived lines and performed RT-PCR in order to generate cDNA. The *ABCA7* cDNA was generated from peripheral blood mononuclear cells (PBMCs), as well as day 23 neuronal precursor cells (NPCs). This cDNA was then amplified across the 44 base pair deletion, which presented with a visually detectable double band in the patient, which is a heterozygote expressing both a wild type and mutated form of the gene (A). For the *SORL1* lines, we were able to determine that the gene was being expressed, but the agarose gel did not permit resolution of the single base pair deletion (B). Finally, we were able to detect via immunocytochemistry (ICC) that day 35 neurons were expressing the ABCA7 protein (C), demonstrating that phenotypes detected at this relatively early time point could be related to disruption of this protein.

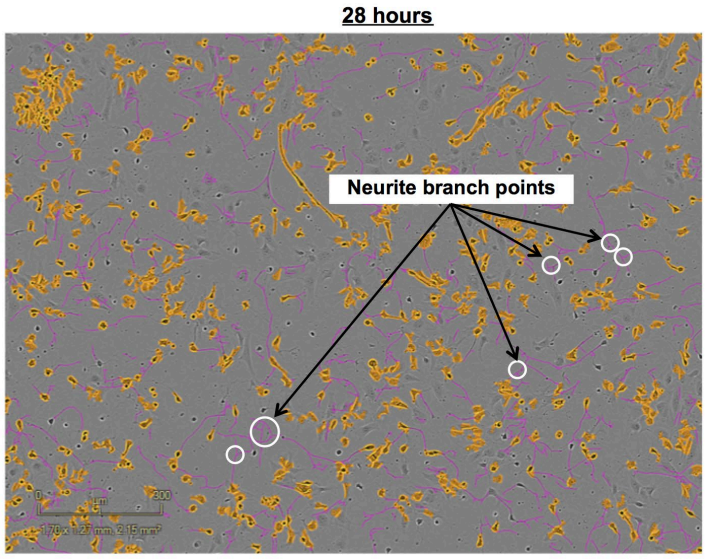
Neurons derived from the *ABCA7* lines were plated at day 35 for ICC and fixed at day 40. Both the control and AD case demonstrated that they were expressing neuronal markers including the broad neuronal marker synapsin as well as the more mature neuronal marker MAP2.



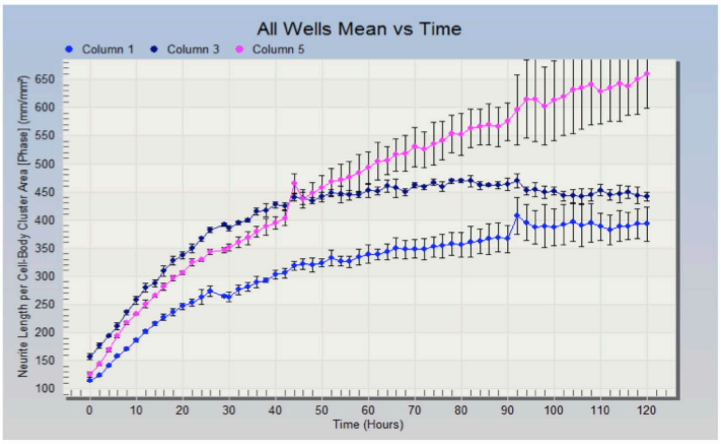
We have optimized cellular function assays that measure beta amyloid, tau, neurite growth, and rate of apoptosis. ELISA assays are used to analyze of the different amyloid beta species (A□40 and A□42) thought to be pathogenic in AD from the cell culture supernatant of the neurons. We have tested one of the *ABCA7* patient lines that carries the 44 base pair deletion in parallel with an ethnically matched (AA) control line from a neurologically normal individual at a relatively early neuronal stage (day 40-45) as well as an older culture from the same lines. At both time points, neurons from the patient line had a higher level of amyloid beta 40 compared to neurons generated from the control individual. Furthermore, amyloid beta is being secreted at higher levels in both lines as the cells age.

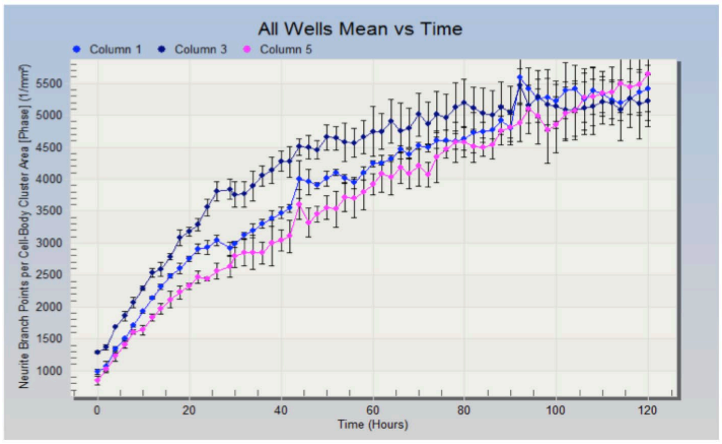
In addition, morphological measures of axon and synapse formation were assessed in differentiating neuronal cultures using the IncuCyte ZOOM live cell imaging system (Essen BioScience). The IncuCyte ZOOM supports high content phase contrast, green fluorescence and red fluorescence imaging modes. The differentiating neurons were plated into triplicate wells of a 24-well plate coated with poly-D-lysine, laminin, and fibronectin. The IncuCyte ZOOM can capture images of growing cultures. The results of 4 fields of view/well from each well were analyzed to measure the cell body area (shown in yellow-orange below) as well as identify the neurites (purple) and the neurite branch points (encircled). The measurements for axon length/cell body area and branch points/cell body area were analyzed by the Neurotrack software.



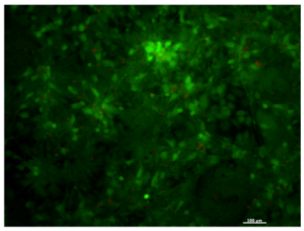


Preliminary results with one *ABCA7* line (shown in royal blue below) compared to two African American control lines lacking the 44 base pair deletion is shown below for days 35-40. The cells were placed into the IncuCyte ZOOM and analyzed for 5 days with imaging occurring every 4 hours. It appears that neurite length in the AD case, when normalized to the cell body area, may be reduced compared to controls (below, left). However, neurite branching does not appear to be different in the case compared to the control lines (below right).

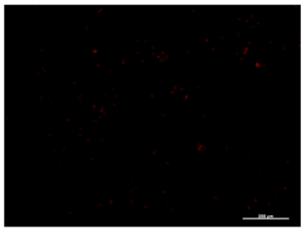




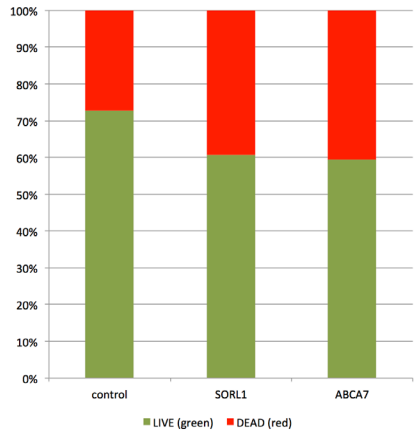
Furthermore, although the most consistent findings in AD-specific iPSC derived neurons has been increased levels of A□42, A□40, and phosphorylated tau, there have also been reports of neurons from AD patient-derived iPSC lines were more susceptible to cell death (Duan, et al, Mol. Neurodegener, 2014). Therefore, we sought to analyze the health and viability of the cultures throughout the differentiation process using the LIVE/DEAD Viability/Cytotoxicity Kit (Life Technologies). Love cells stain green, while dead cells are marked red. Initial experiments performed on day 35 cells under the normal growth conditions demonstrated that there may be an increase in cell death in neurons generated from *ABCA7* and *SORL1* AD patients. In the future, we may modify the assay to further treat the cells by adding deregulated calcium (treatment with ionomycin) or glutamate exposure (excitotoxicity) to the cultures.



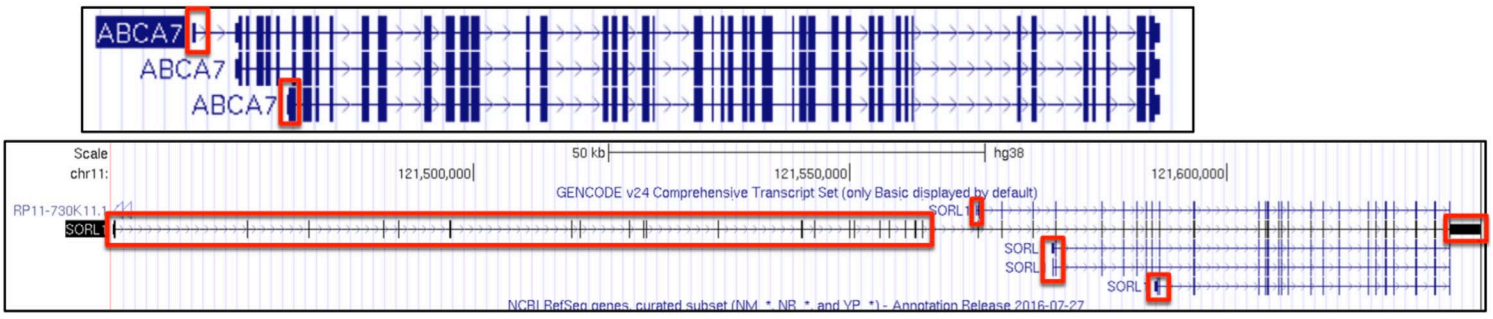
LIVE = **green**
Calcein AM



DEAD = **red**
Ethidium homodimer-1

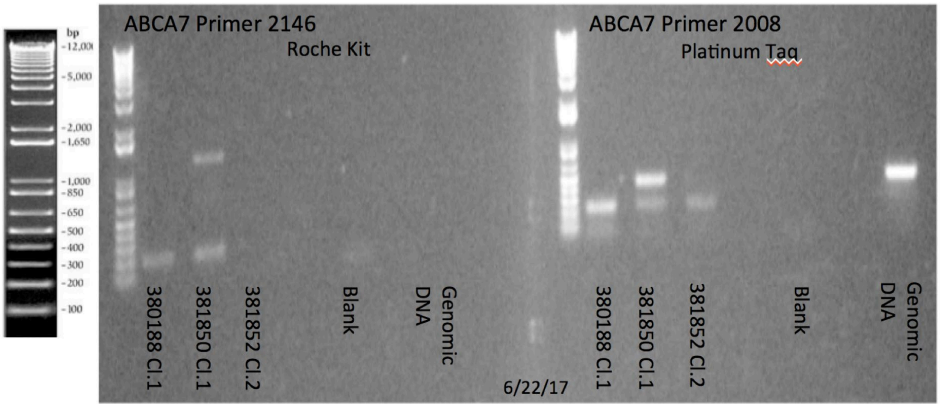


Our initial evaluation of iPSCs and cells being differentiated into neurons demonstrated that both *ABCA7* and *SORL1* were being expressed. However, each of these genes has multiple isoforms that could potentially be expressed. In order to best predict the size the protein that might be generated from the full length and potentially truncated versions of these genes, we designed isoform specific primers: 2 primer sets for *ABCA7* and 5 primer sets for *SORL1*. Primers were designed that only hybridized to the unique regions of these genes (boxed in red below) for 2 isoforms of *ABCA7* and five isoforms of *SORL1* in order to perform reverse transcription polymerase chain reaction (RT-PCR).



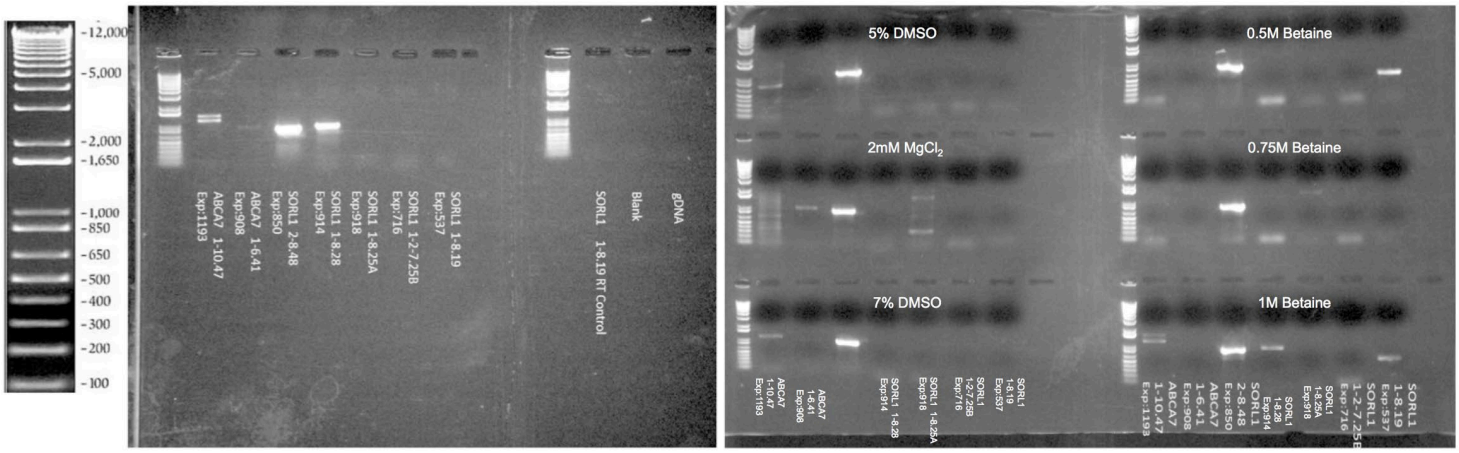
The gel below shows the 5 primer sets for the *SORL1* gene being tested in cDNA generated from RNA isolated from iPSC and day 35 neurons in both AD case and control lines. The primers are named by the size of the corresponding protein that the specific isoform would encode. The results below demonstrate that there appears to be bands for some of the samples for the isoforms that encode proteins of 829, 1124, 1158, and 2214 amino acids. The isoform that encodes that largest protein with 2214 amino acids, appears most consistently across the 4 different lines in both iPSC and day 35 neurons.

For the *ABCA7* isoforms, we were able to specific primers for 2 of the 3 recognized isoforms. The gel below shows that both isoforms were recognized in cDNA isolated from RNA from day 35 neurons in lines derived from both case and control individuals.

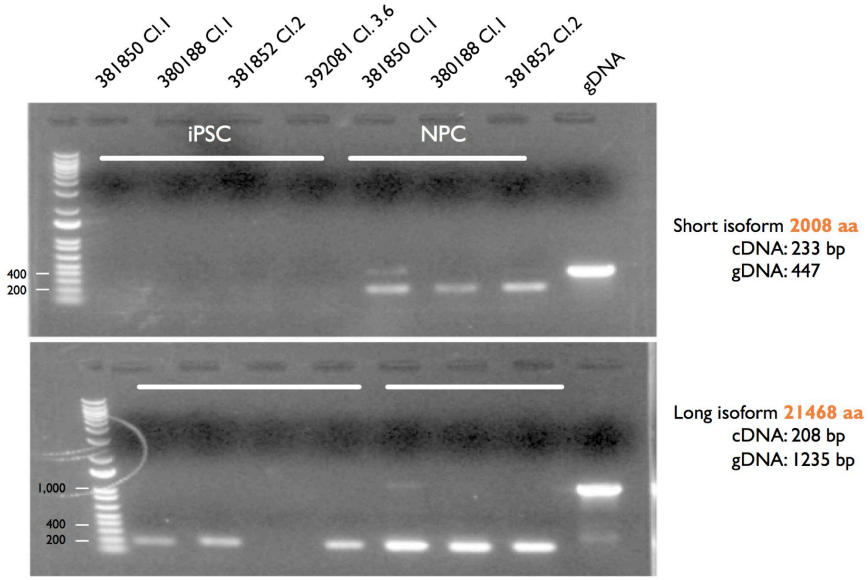


In-Silico:
Primer 2146: 208bp
Primer 2008: 233bp

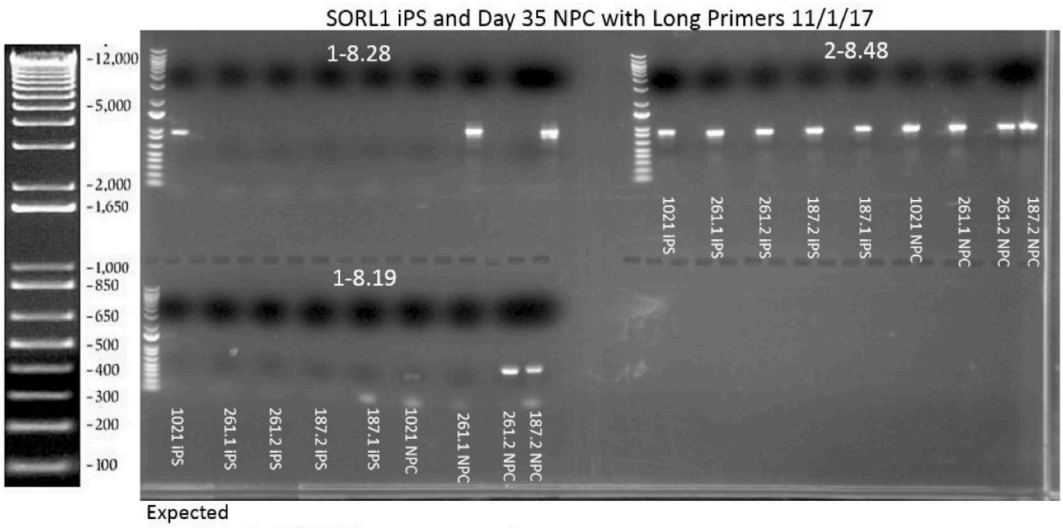
Our initial experiments had bands demonstrating that there was contamination of genomic DNA. We therefore redesigned new primers that would not allow for genomic DNA to be amplified under standard touchdown PCR conditions. These primers are still being optimized (see below).



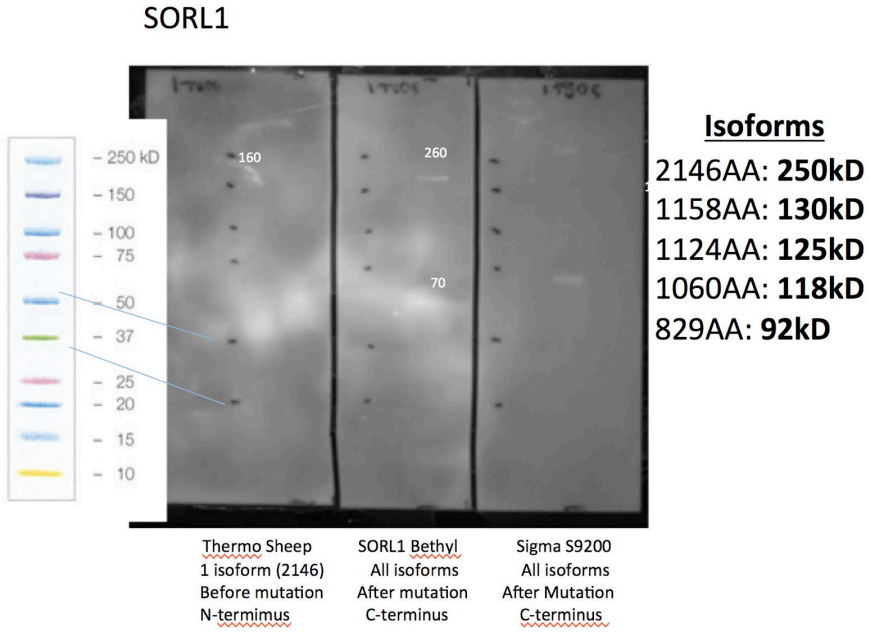
Results confirmed that both of the *ABCA7* isoforms are being expressed, with the longer isoform being expressed in most of the cell lines at both the stem cell stage as well as at the neuronal precursor cell (NPC) stage.



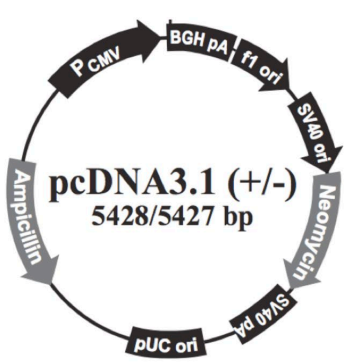
For the *SORL1* isoforms, we were able to confirm that three were being expressed; one of these isoforms was recognized to be present in all cell lines at both the stem cell and NPC stage, while the other isoforms were only found in some cell lines at distinct time points. These results show that we are capturing time points where our gene of interest is being expressed and that we can potentially identify phenotypic distinctions in the iPSC and NPCs in patient-derived iPSC with the variants of interest in *ABCA7* and *SORL1*.

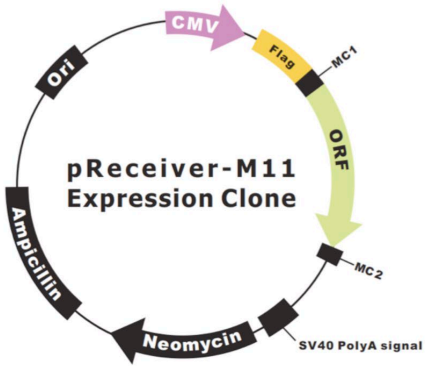


Since we have established some of the isoforms that are being expressed in both the iPSC and the young developing neurons, we are in the process of optimizing western blot protocols in order to determine whether the individuals with the deletions in *ABCA7* and *SORL1* are generating truncated versions of the protein, which may interfere with function through dominant negative mechanisms, or if no truncated protein is produced and that disease might be a consequence of loss-of-function. Both of the largest isoforms of *ABCA7* and *SORL1* were confirmed to be generated and these are predicted to produce large proteins ~250 kDa in size. We have purchased antibodies that detect each of the proteins, which antibodies located both upstream and downstream of the deletions. While we are attempting to optimize the western protocol for large proteins by using lower percentage (7%) Tris-Acetate gels, using PVDF membranes instead of nitrocellulous and transferring for an extended period of time overnight in the cold room. While staining with Ponceau S demonstrates the successful transfer of proteins of a large size, the antibodies have been giving ambiguous results thus far (see below).

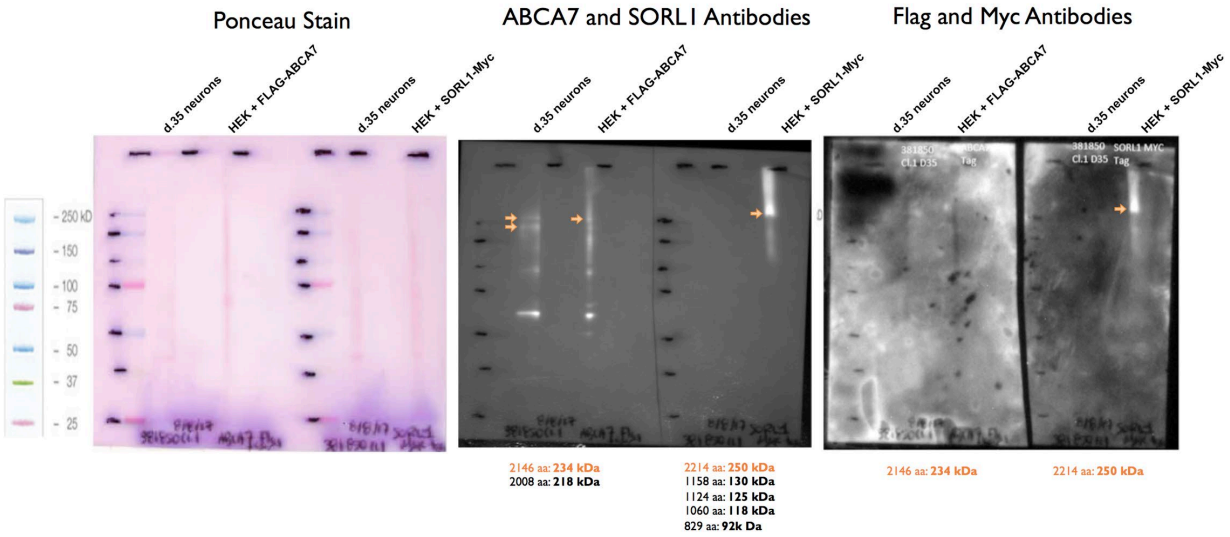


To ensure that the bands that are being identified on the western blot are specific to the protein of interest, we obtained plasmids with tagged versions of the genes of interest: SORL-MYC pcDNA3.1 (Bohm, et al, JBC, 2006, obtained from the St. George-Hyslop lab) and FLAG-ABCA7 in the pReceiver-M11 plasmid from GeneCopoeia (see figure below).

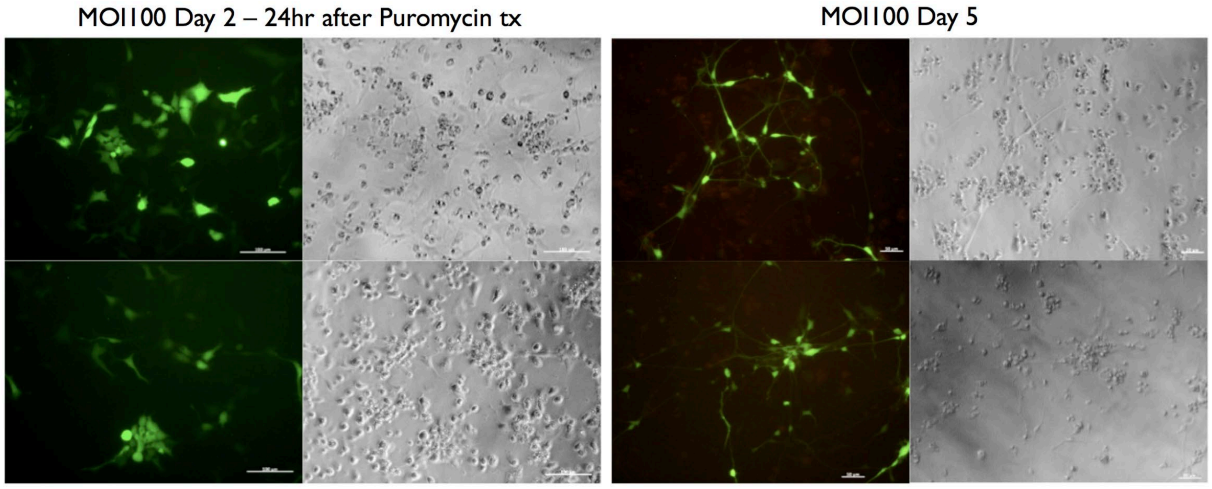




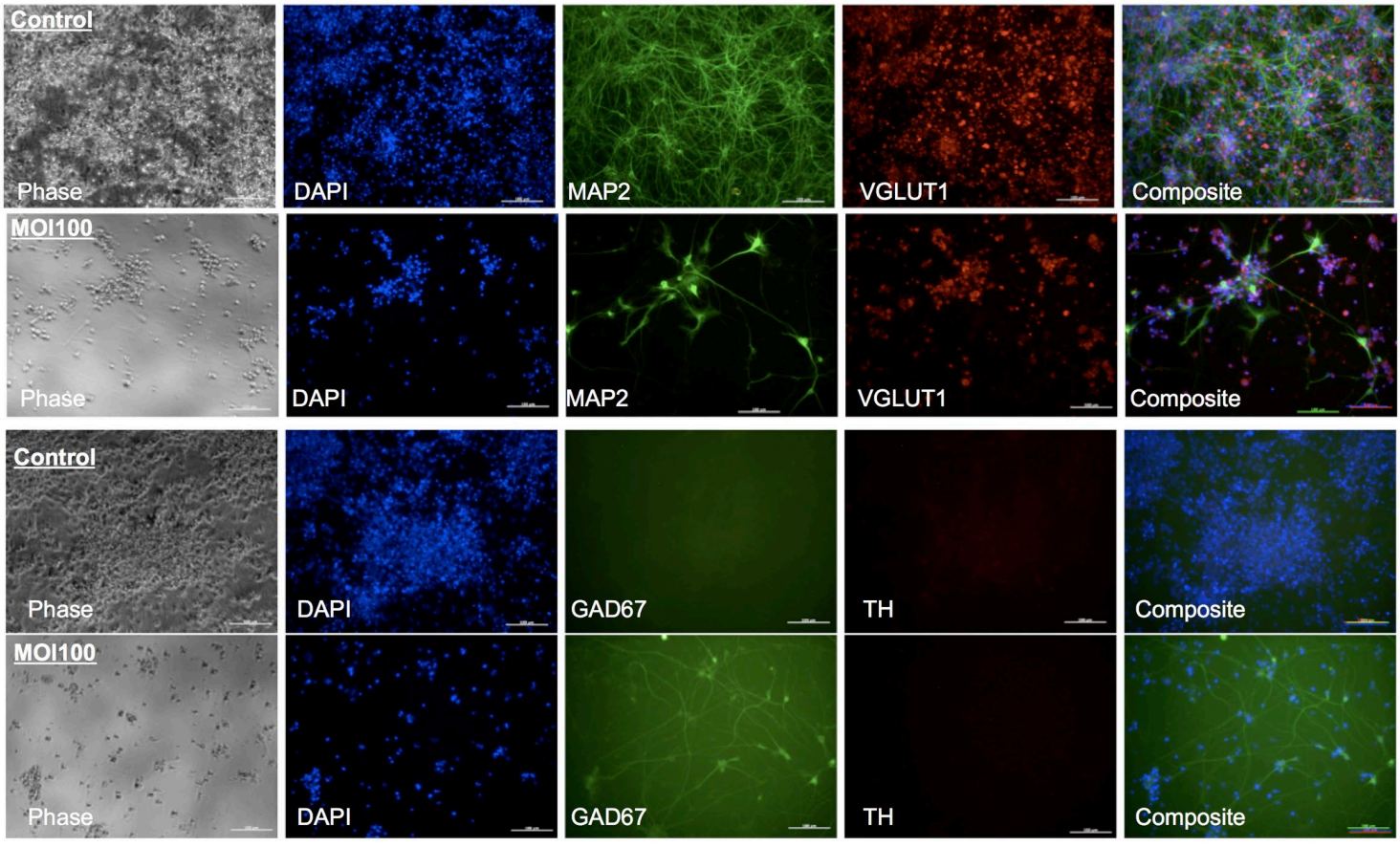
Both of these plasmids were transfected into HEK293 cells with jetPRIME and cells collected after 72 hours. The protein was extracted and run alongside of protein extracted from day 35 neurons (figure below). The Ponceu stain demonstrated that there was protein being transferred to the membrane. The blots were initially probed with antibodies for ABCA7 and SORL1, and then stripped and immunoblots performed for Flag and Myc, respectively. The results demonstrated that the FLAG-ABCA7 protein appears to be getting the same cross-reactivity as seen in the day 35 neurons extract, but no signal from the FLAG antibody. In contrast, the SORL-Myc protein was successfully detected by both the SORL1 and Myc antibodies, although the day 35 neurons did not appear to be expressing any SORL1 protein. We will continue to work with these tools in order to resolve the question as to whether or not ABCA7 and SORL1 are being expressed in neurons.



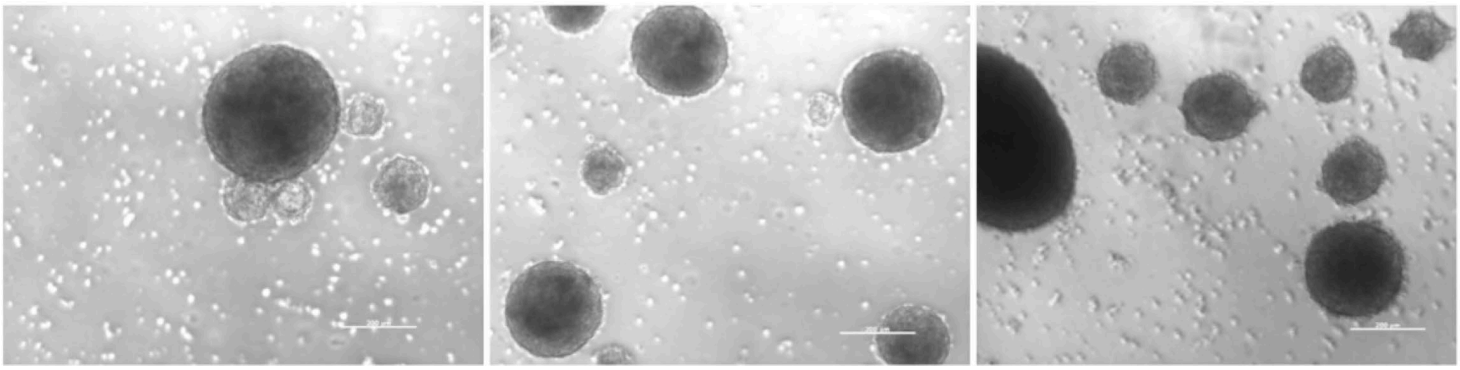
In addition, we have been optimizing a protocol to obtain more homogenous cultures of excitatory neurons through the viral transduction of *Neurogenin 2* (*NGN2*) onto neuronal progenitor cells (NPCs, Ho, et al, *Methods*, 2016). This is performed using viral particles containing TetO-mNgn2-P2A-Puro. There was also a doxycycline dependent expression of GFP, allowing the live cell cultures to be observed for adequate transduction (see below). 2 days following transduction, puromycin selection was also used to eliminate cells that were not transduced. 1 week following transduction, the cells were treated to Ara-C, which reduces the proliferation of non-neuronal cells.



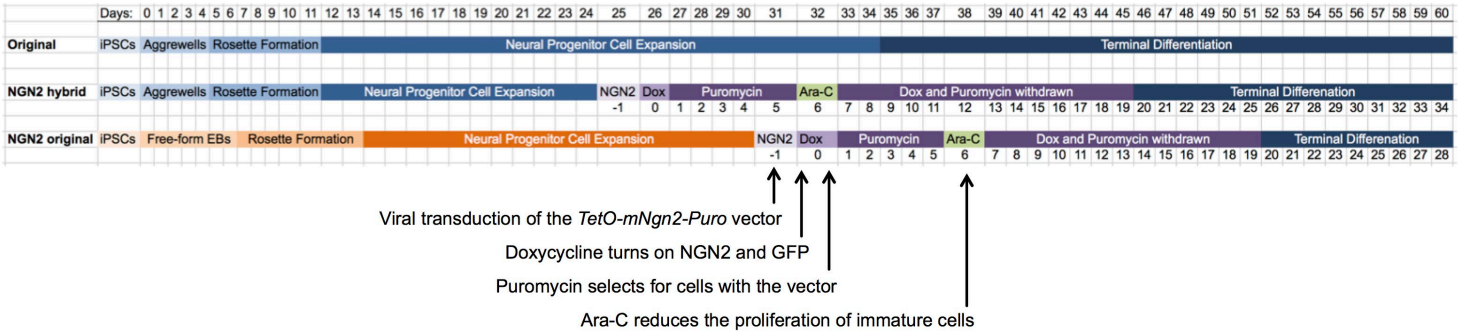
At 20 days following transduction, immunocytochemistry (ICC) was performed on the cells to determine the composition of the cultures (see below). This experiment demonstrated an acceleration of the rate of neuronal maturation between the cells that were not transduced with *Ngn2* (control) and those cells that were transduced at a viral MOI of 100. However, we are further optimizing the protocol by doing the following: 1.Increasing the viral load to the cells, 2. Allowing the cells to mature longer, past the initial 20 day time point that was assessed, 3. Restarting the experiment from an earlier time point, using induced pluripotent stem cells (iPSC) as a starting point and following the protocol outlined in Topol, et al, *Journal of Visualized Experiments*, 2015 in order to follow the differentiation protocol more as it was initially published.



The protocol from Topol, et al, allows embroid bodies to be formed spontaneously, as opposed to using an Aggrewell plate, and are therefore are of different sizes (see below). These will then be fed small molecules to push them to a neuronal fate and develop into neural rosettes and then neural progenitor cells prior to being transduced with *Ngn2*.

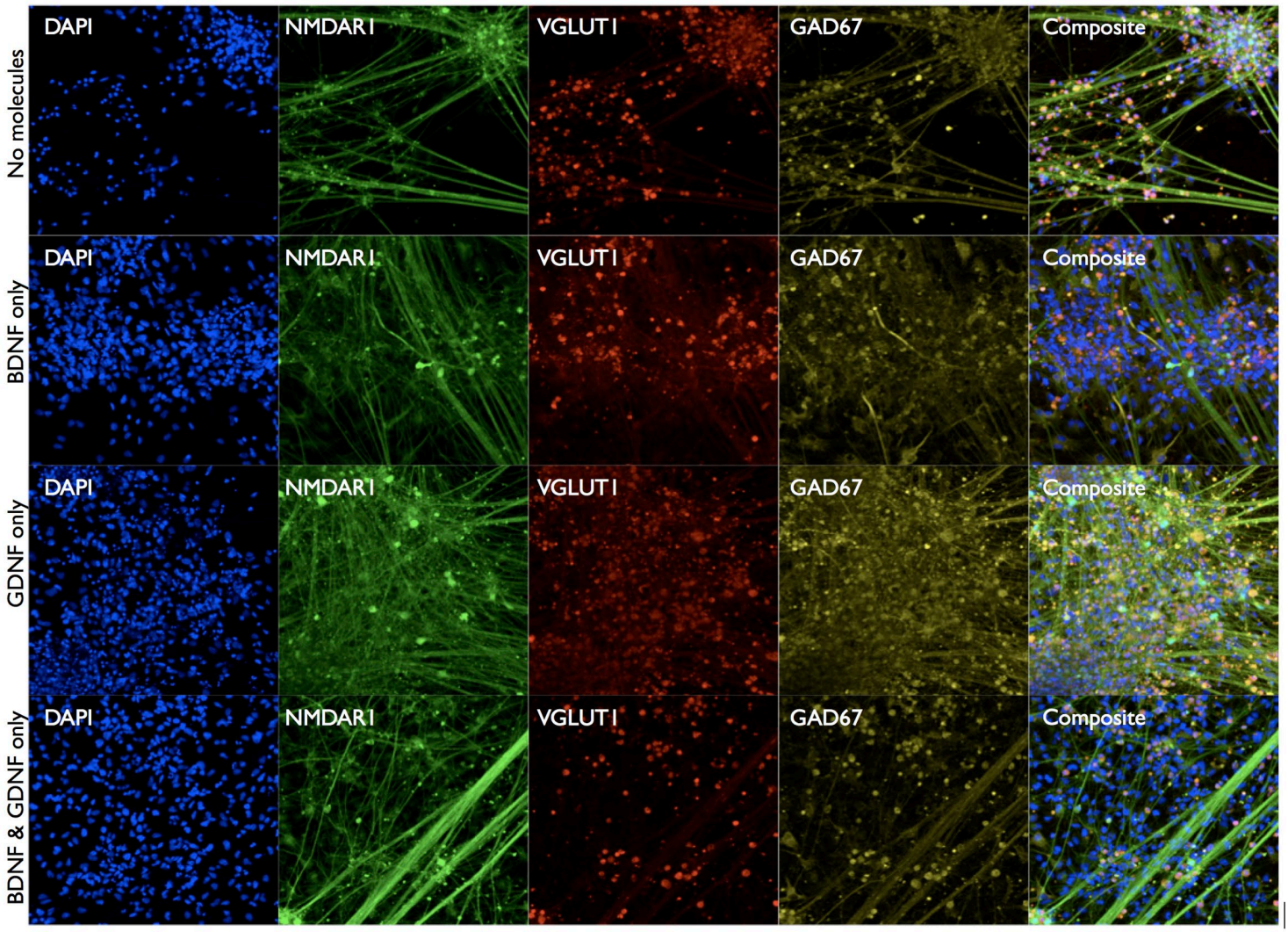


We continued optimizing a protocol to obtain more homogenous cultures of excitatory neurons through the viral transduction of *Neurogenin 2* (*NGN2*) onto neuronal progenitor cells (NPCs, Ho, et al, *Methods*, 2016). This is performed using viral particles containing TetO-mNgn2-P2A-Puro. There was also a doxycycline dependent expression of GFP, allowing the live cell cultures to be observed for adequate transduction. 2 days following transduction, puromycin selection was also used to eliminate cells that were not transduced. 1 week following transduction, the cells were treated to Ara-C, which reduces the proliferation of non-neuronal cells. Below are the timelines for our original differentiation method (no neurogenin used), a hybrid method that transduces our neuronal precursor cells (NPCs) with neurogenin and, lastly, the neurogenin protocol using NPCs generated as outlined in Topol, et al, *Journal of Visualized Experiments*, 2015.

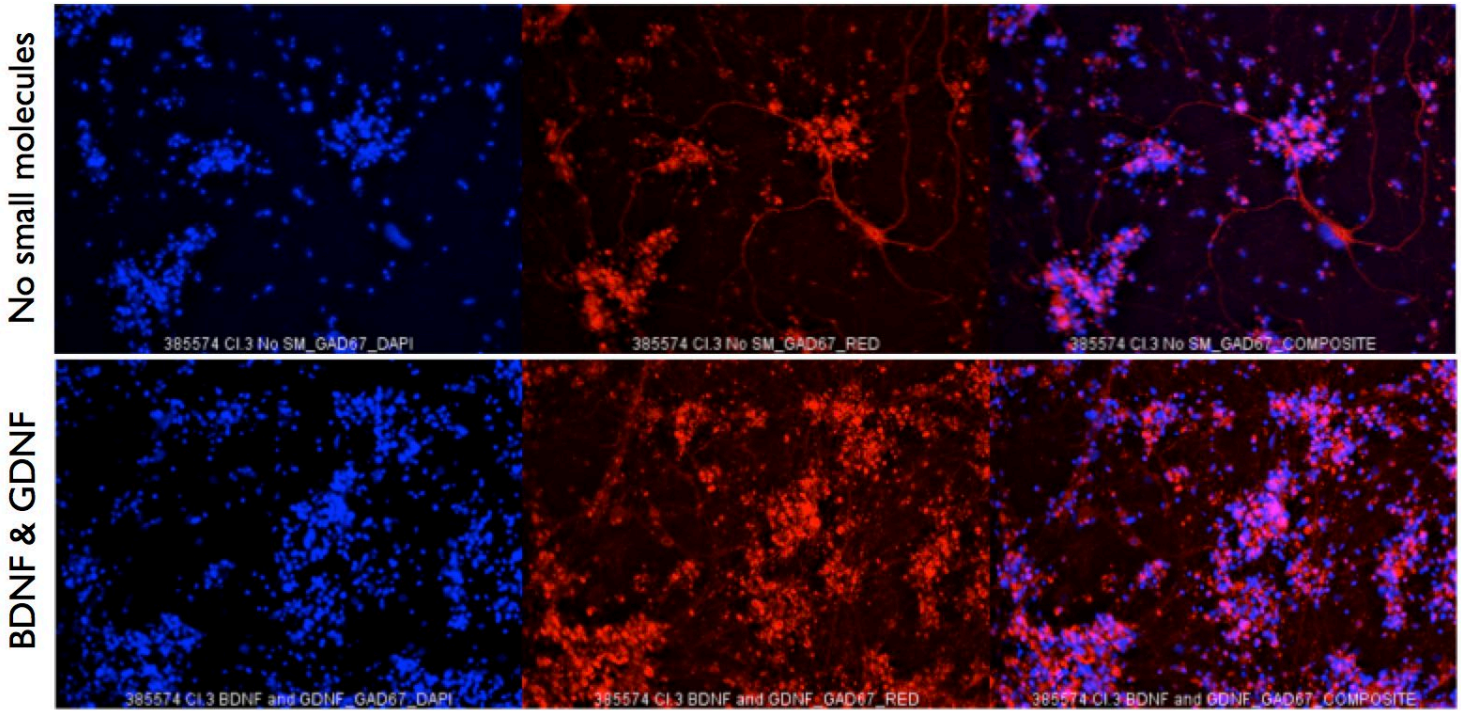


Our first attempt used the NGN2 hybrid method and produced neurons and removed glia cells from the culture. However, we were expecting to see glutamatergic neurons, yet found the neurons stained positive for GAD67, a GABAergic marker. This was contradicting the results that were initially published. We therefore set out to modify our initial attempt of this protocol using two distinct variations: 2. Using the hybrid method and modifying whether the additives BDNF or GDNF are added, which may be modifying whether the cells are going to an excitatory or inhibitory cell fate. 2. Restarting the experiment from an earlier time point, using induced pluripotent stem cells (iPSC) as a starting point and following the protocol as outlined in Topol, et al, *Journal of Visualized Experiments*, 2015 in order to follow the differentiation protocol more as it was initially published using BDNF and GDNF or no small molecules.

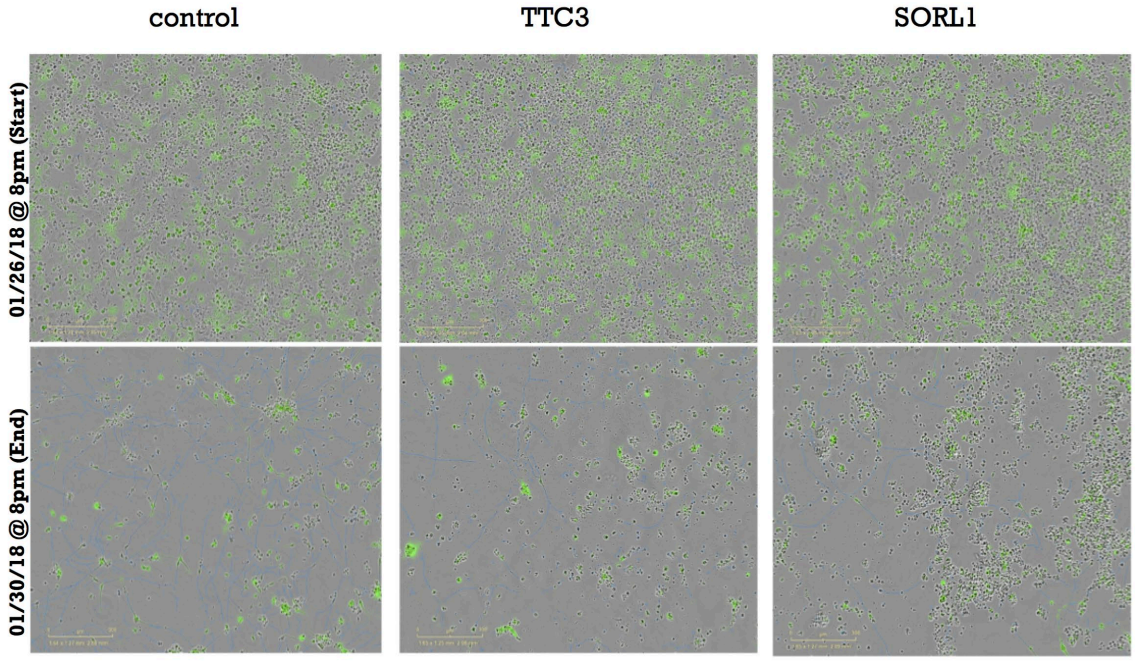
When we tried to modify the BDNF and GDNF molecules, we did not find a difference in the immunofluorescence of either VGLUT1, a vesicular glutamate transporter (shown in red in the images below) or GAD67, a glutamic acid decarboxylase that catalyzes the conversion of L-glutamic acid to the inhibitory GABA (shown in yellow in the images below).



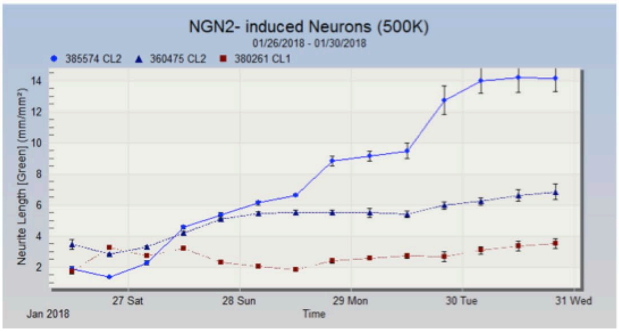
We also replicated the experiment from an earlier time point, using induced pluripotent stem cells (iPSC) as a starting point and following the protocol as outlined in Topol, et al, *Journal of Visualized Experiments*, 2015 in order to follow the differentiation protocol more as it was initially published using BDNF and GDNF or no small molecules. However, staining for GAD67 was still positive, signifying the presence of GABAergic neurons.

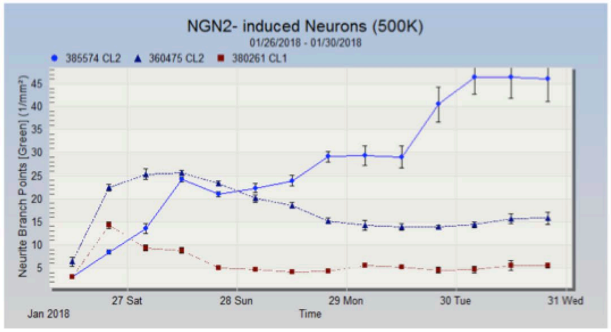


Given the results above, we have decided to move forward with the hybrid NGN2 method, where the neural rosettes will be uniform from being grown in a aggrewell dish, but the neurons will mature more rapidly.
We proceeded with the NGN2 hybrid protocol shown above with a SORL1 AD patient iPSC line, a TTC3 AD patient iPSC line, and a control iPSC line. These lines were placed in the Incucyte Zoom machine in order to record how the morphology changed over time. In the initial days, there was cell death see as only cells with the NGN2 vector and that were mature were able to survive the puromycin and Ara-C selections (see figure below). Cell bodies are distinguished in green, while neurite outgrowths are shown in blue.

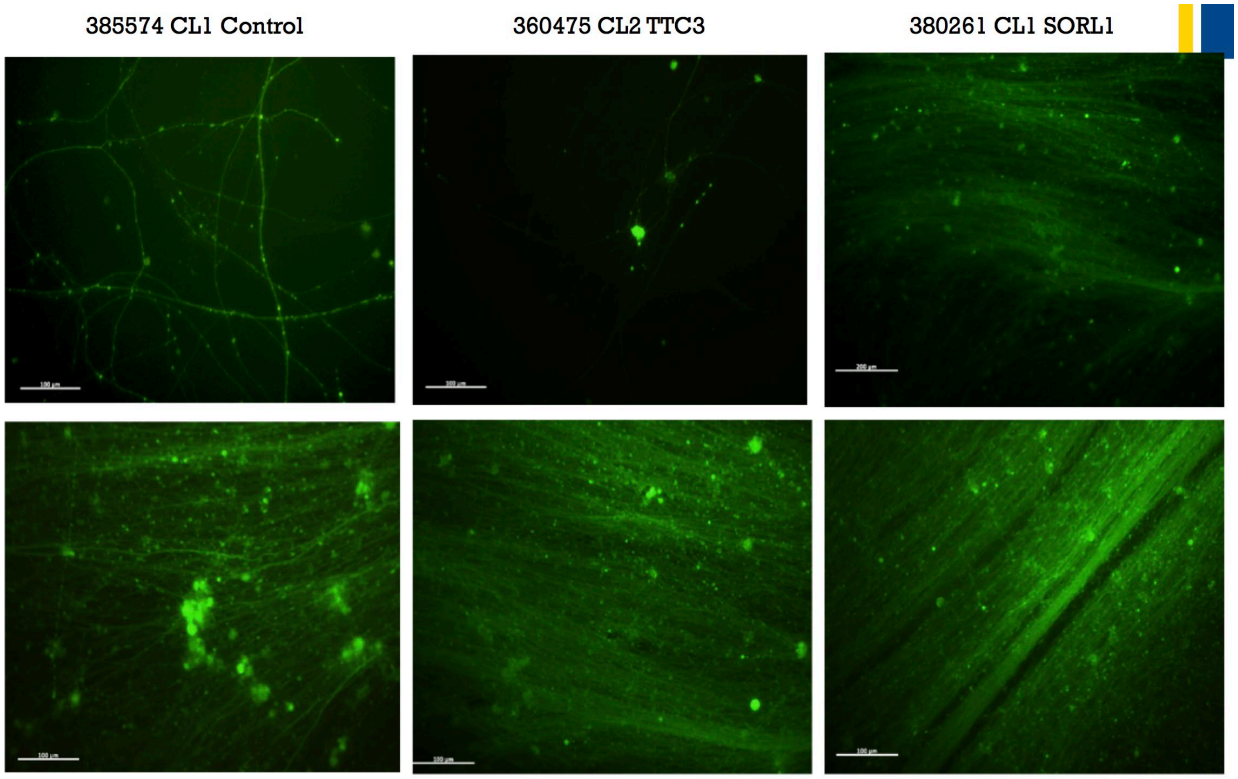


Preliminary results show that there may be better growth in the control line (blue circle on the graphs below) compared to the two AD lines.





We also grew the cultures longer and found that the neurons in large vessels (6 well dishes) were able to develop quite extensive and dense neurite outgrowths (see figure below). These cultures will be utilized in order to collect neuronal cells and evaluate cellular and AD specific phenotypes.



KEY RESEARCH ACCOMPLISHMENTS:

2015

- Variant calling and quality control processing of these samples completed on 55 Hispanic individuals submitted by Columbia and 51 NH-White samples from the University of Miami and Vanderbilt University.
- Analysis (variant annotation and filtering) completed on samples of 55 Hispanic individuals submitted by Columbia and 51 NH-White samples from the University of Miami and Vanderbilt University.
- Identity-by-descent analysis of Hispanic families is complete.
- Identification of 125 top candidate variants for follow-up genotyping is complete.
- Genotyping of 125 top candidate variants in the Hispanic families and a cohort of 500 Hispanic cases controls is complete.
- Analysis of the 125 top candidate variants in the Hispanic families and a cohort of 500 Hispanic cases and controls is complete, with 20 top candidates identified for follow-up, including a 44 base-pair deletion in the known LOAD gene *ABCA7*.
- Follow-up of these 20 top candidates in large cohorts of Hispanic, AA and NHW AD cases and controls helped prioritize the top candidate genes and confirmed the association of the deletion in *ABCA7* to increased risk of AD.
- Identification of rare coding variants in *SORL1, PSEN1, and MAPT* in EOAD cases including a potential link between SORL1 and Parkinsonism.
- Identification of 5 candidate early-onset Alzheimer disease genes (*HSPG2, DOCK3, OGT, CLSTN1,* and *PARK2*) through identification of NHW EOAD cases with shared rare coding variants with damaging potential in genes interacting with known EOAD genes.

2016

- Publication of the link between a 44 base-pair deletion in the known LOAD gene *ABCA7* as a risk factor for Alzheimer's in both Hispanics and African Americans.
- Submittal of a manuscript describing a link between SORL1 and Parkinsonism in Alzheimer's cases.
- Sanger confirmation and genotyping in cases and controls of variants in 5 genes that interact with known EOAD genes and have shared rare coding variants with damaging potential in 2 or more NHW EOAD was completed. No significant variant was identified.
- Analysis of EOAD exome chip association data and comparison to WES sequencing identifies several candidate genes for EOAD, including the endocytosis related gene *PSD2*.
- Ascertainment of patient samples for iPSC derivation from the Hussman Institute Human Genomics (HIHG) cohort and the Columbia University cohort with genetic variants in SORL1, SEC16A, TTC3, and ABCA7.
- PBMCs reprogrammed into iPSC from the patient samples and validated for pluripotency and karyotype.
- Optimization of assay for pathogenic beta species from the culture supernatant and pathogenic tau from lysates from iPSC-derived neurons.

2017

- Publication of the manuscript describing the *SORL1* alterations that we identified in EOAD and LOAD families, and how some of these individuals exhibit parkonsonian features (Cuccaro, et al, 2016).
- Manuscript describing the results from the whole exome sequencing is currently under revision following initial reviewer comments from *JAMA Neurology* (Kunkle, et al).
- iPSC lines have been generated from 2 AD individuals with the *SORL1* nonsense change, 2 AD individuals with the *ABCA7* 44 base pair deletion (rs142076058), and a single AD individual with the *SEC16A* alteration.
- Confirmed that *ABCA7* RNA and protein is being made in the iPSC and neuronal precursor cells
- Verified that *SORL1* RNA is being expressed in the iPSC and day 35 neurons
- Optimized the following functional assays:
    - Amyloid beta quantification of pathogenic species in the culture supernatant via ELISA
    - Live measuring of growing neurons and measuring neurons development through the neurite length and number of branch points (a sign of neuron maturity)

    - LIVE/DEAD apoptosis staining through immunocytochemistry
- Initial results indicate that the *ABCA7* deletion carrying patients may have decrease in neuronal maturity, as measured by shorter neurite length in day 35-40 neurons compared to controls.
- Preliminary data shows that both the ABCA7 and SORL1 alterations may make the cells more vulnerable to cell death, as shown through the LIVE/DEAD assay.
- A R01 grant entitled "*Genetic Epidemiology of Early-Onset Alzheimer's disease in Caribbean Hispanics and non-Hispanic Whites*" was awarded to Drs. Beecham (University of Miami) and Reitz (Columbia University). This grant will utilize whole genome sequencing of early onset AD families in a set of multiplex families using extreme phenotyping designs to increase statistical power by creating more homogeneous and genetically loaded populations, leading to the potential to reveal genetic risk factors and mechanisms difficult to identify in more heterogeneous datasets.

2018

- Comparison of the top 19 Hispanic candidates from the follow-up genotyping to the Caucasian EOAD WES samples is ongoing.
  Analysis of candidate variants/loci in our large LOAD case control data set is ongoing.
- Publication of the manuscript describing the *SORL1* alterations that we identified in EOAD and LOAD families, and how some of these individuals exhibit parkonsonian features (Cuccaro, et al, 2016).
- Manuscript describing the results from the whole exome sequencing is currently under revision following initial reviewer comments from *JAMA Neurology* (Kunkle, et al).
- iPSC lines have been generated from 2 AD individuals with the *SORL1* nonsense change, 2 AD individuals with the *ABCA7* 44 base pair deletion (rs142076058), and a single AD individual with the *SEC16A* alteration.
- Confirmed that *ABCA7* RNA and protein is being made in the iPSC and neuronal precursor cells
- Verified that *SORL1* RNA is being expressed in the iPSC and day 35 neurons
- Optimized the following functional assays:
    - Amyloid beta quantification of pathogenic species in the culture supernatant via ELISA
    - Live measuring of growing neurons and measuring neurons development through the neurite length and number of branch points (a sign of neuron maturity)
    - LIVE/DEAD apoptosis staining through immunocytochemistry
- Initial results indicate that the *ABCA7* deletion carrying patients may have decrease in neuronal maturity, as measured by shorter neurite length in day 35-40 neurons compared to controls.
- Preliminary data shows that both the ABCA7 and SORL1 alterations may make the cells more vulnerable to cell death, as shown through the LIVE/DEAD assay.
- A R01 grant entitled "*Genetic Epidemiology of Early-Onset Alzheimer's disease in Caribbean Hispanics and non-Hispanic Whites*" was awarded to Drs. Beecham (University of Miami) and Reitz (Columbia University). This grant will utilize whole genome sequencing of early onset AD families in a set of multiplex families using extreme phenotyping designs to increase statistical power by creating more homogeneous and genetically loaded populations, leading to the potential to reveal genetic risk factors and mechanisms difficult to identify in more heterogeneous datasets.
- Isoform specific primers were designed for *SORL1* and *ABCA7*. Results show that 4 of the *SORL1* isoforms and 2 of the 3 known ABCA& isoforms are being expressed in day 35 neurons
- Western blot protocols are being optimized for these large proteins encoded by *SORL1* and *ABCA7* to determine whether a truncated version of either protein is being generated in the individuals that carry the alterations that contribute to AD risk. To facilitate this, two plasmids with tagged versions of these genes have been acquired to act as positive controls.
- Designed primers to differentiate 5 distinct isoforms for *SORL1* and 2 of the 3 distinct isoforms of *ABCA7*. It appears that 4 of the *SORL1* isoforms are being expressed in day 35 neurons and 2 of the 3 *ABCA7* isoforms was identified. We are in the process of optimizing the western blot protocols for these large proteins. In addition, we have acquired 2 plasmids with tagged versions of these genes in order to act as positive controls in the western blot experiments.
- We have been optimizing a protocol to obtain more homogenous cultures of excitatory neurons through the viral transduction of *Neurogenin 2* (*NGN2*, Ho, et al, 2016). Initial experiments demonstrated an acceleration of the rate of neuronal maturation.
- We are optimizing RT-PCR and western blots to test *SORL1* and *ABCA7* expression in neurons.
- Hurricane Irma made landfall in Florida in September of 2017, causing the University of Miami Miller

School of Medicine to close for about a week. Experiments were resumed as soon as possible following the storm, but there were some difficulties including limited public transportation due to blocked roads and debris, and some personnel evacuating to distant locations.

- We have been optimizing a protocol to obtain more homogenous cultures of excitatory neurons through the viral transduction of *Neurogenin 2* (*NGN2*, Ho, et al, 2016). Experiments to modify either small molecules (BDNF/GDNF) and to change the manner in which the neuronal precursor cells were generated were attempted to optimize conditions.
- We found that multiple isoforms of *SORL1* and *ABCA7* are expressed in both stem cells and young neurons.
- The manuscript titled “Whole exome sequencing of early-onset Alzheimer’s disease highlights candidate risk genes involved in endo-lysosomal transport” by Kunkle, et al, has published in JAMA Neurology. Epub Jul 24 (PMID: 28738127).
- A Florida Department of Health, Ed and Ethel Moore AD Research Program grant titled “The Role of *SORL1* in Alzheimer’s Disease Pathogenesis” was awarded to Drs. Dykxhoorn and Cukier. This grant will focus on severe truncating SORL1 alteration in a multiplex family and use stem cells models and CRISPR genome editing to determine the importance of the gene to cellular phenotypes.
- A BrightFocus Alzheimer’s Disease Research grant titled “Elucidating the Cell-Specific Roles of *ABCA7*” was awarded to Drs. Cukier and Dykxhoorn. This grant will evaluate the effects of an African-specific deletion in *ABCA7* that is associated with Alzheimer’s disease in both neurons and microglia generated from patient-derived stem cell lines, and determine whether one cell type is more vulnerable to *ABCA7* dysfunction.

REPORTABLE OUTCOMES:

Manuscripts:

Cukier HN*, Kunkle BW*, Vardarajan BN*, Rolati S, Hamilton-Nelson KL, Kohli MA, Whitehead PL, Dombroski BA, Van Booven DJ, Lang R, Dykxhoorn DM, Farrer LA, Cuccaro ML, Vance JM, Gilbert JR, Beecham GW, Martin ER, Carney RM, Mayeux R, Schellenberg GD, Byrd GS, Haines JL, Pericak-Vance MA, Alzheimer's Disease Genetics Consortium. ABCA7 Frameshift Deletion Associated with Alzheimer's Disease in African Americans. Neurology: Genetics, 2016 June 2(3):e79. PMCID: PMC4871806. (Appendix I)

Cuccaro ML, Carney RM, Zhang Y, Bohm C, Kunkle BW, Vardarajan BN, Whitehead PL, Cukier HN, Mayeux R, St George-Hyslop P, Pericak-Vance MA. *SORL1* mutations in early- and late-onset Alzheimer disease. Published in Neurol Genet. 2016 Oct 26;2(6):e116. (Appendix II)

Kunkle BW, Carney RM, Kohli MA, Naj AC, Hamilton-Nelson KL, Whitehead PL, Wang L, Lang R, Cuccaro ML, Vance JM, Byrd GS, Beecham GW, Gilbert JR, Martin ER, Haines JL, Pericak-Vance MA. Targeted sequencing of ABCA7 identifies splicing, stop-gain and intronic risk variants for Alzheimer disease. Neurosci Lett. 2017 May 10;649:124-129. (Appendix III)

Cukier HN, Kunkle BK, Hamilton KL, Rolati S, Kohli MA, Whitehead PL, Jaworski J, Vance JM, Cuccaro ML, Carney RM, Gilbert JR, Farrer LA, Martin ER, Beecham GW, Haines JL, Pericak-Vance MA. Exome Sequencing of Extended Families with Alzheimer's Disease Identifies Novel Genes Implicated in Cell Immunity and Neuronal Function. J Alzheimers Dis Parkinsonism. 2017 Aug;7(4). pii: 355. (Appendix IV)

Kunkle BW, Vardarajan BN, Naj AC, Whitehead PL, Rolati S, Slifer S, Carney RM, Cuccaro ML, Vance JM, Gilbert JR, Wang LS, Farrer LA, Reitz C, Haines JL, Beecham GW, Martin ER, Schellenberg GD, Mayeux RP, Pericak-Vance MA. Early-Onset Alzheimer Disease and Candidate Risk Genes Involved in Endolysosomal Transport. JAMA Neurol. 2017 Sep 1;74(9):1113-1122. (Appendix V)

Whole-exome sequencing of Hispanic families identifies novel candidate genes for early-onset Alzheimer disease. *In preparation.*

Presentations:

American Society for Human Genetics (ASHG), Boston, MA, October 22-26, 2013:
Kunkle BW, Kohli MA, Vardarajan BN, Reitz C, Naj AC, Whitehead PL, Martin ER, Beecham GW, Gilbert JR, Farrer LA, Haines JL, Schellenberg GD, Mayeux RP, Pericak-Vance MA, Alzheimer's Disease Genetics Consortium. Whole-exome sequencing in early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes and processes. Platform presentation.

The American Academy of Neurology (AAN) 66th Annual Meeting, Philadelphia, PA, April 26-May 3, 2014:
Reitz C, Kunkle BW, Vandarajan BN, Kohli MA, Naj AC, Whitehead PL, Perry WR, Martin ER, Beecham GW, Gilbert JR, Farrer LA, Haines JL, Schellenberg GD, Pericak-Vance MA, Mayeux RP, Alzheimer's Disease Genetics Consortium. Whole-exome sequencing of Hispanic early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes. Platform presentation.

12th Annual Conference on Alzheimer's Disease and Parkinson's Diseases (AD/PD), Nice, FRA, March 18-22, 2015:
Margaret A. Pericak-Vance, Brian W. Kunkle, Badri Vardarajan, Patrice L. Whitehead, Sophie Rolati, Eden R. Martin, John R. Gilbert, Gary W. Beecham, Richard P. Mayeux, Jonathan L. Haines. Whole-exome sequencing in early-onset Alzheimer disease cases identifies several novel candidate genes. Poster presentation

Carney RM, et al. Novel and known mutations in SORL1, PSEN1, and PSEN2 genes are found in multiplex Alzheimer's disease families with varying age of onset and pathological presentations. Poster presentation

Alzheimer's Association International Conference (AAIC), Copenhagen, Denmark July 12-17, 2014:
Margaret Pericak-Vance, Christiane Reitz, Brian W. Kunkle, Badri N. Vardarajan, Martin A. Kohli, Adam C. Naj, Patrice L. Whitehead, William Perry, Eden Martin, Gary Beecham, John Gilbert, Lindsay A. Farrer, Jonathan

Haines, Gerard D. Schellenberg, Richard Mayeux. Whole-exome sequencing of Hispanic early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer's- related genes. Poster presentation.

Alzheimer's Association International Conference (AAIC), Washington, D.C., July 18-23, 2015:
Gary W. Beecham, PhD; Brian W. Kunkle, PhD, MPH; Badri Vardarajan, PhD; Patrice L. Whitehead, BS; Sophie Rolati, MS; Eden R. Martin, PhD; John R. Gilbert, PhD. Whole-Exome Sequencing in Early-Onset Alzheimer Disease Cases Identifies Novel Candidate Genes. Platform presentation.

Holly N. Cukier, PhD; Brian W. Kunkle, PhD, MPH; Sophie Rolati, MS; Kara L. Hamilton-Nelson, MPH; Martin A. Kohli, PhD; Beth A. Dombroski, PhD; Badri N. Vardarajan, PhD; Patrice L. Whitehead, BS; Derek J. Van Booven, BS; Eden R. Martin, PhD; Gary W. Beecham, PhD; Lindsay A. Farrer, PhD; Michael L. Cuccaro, PhD; Jeffery M. Vance, MD, PhD; Richard Mayeux, MD, MSc; John R. Gilbert, PhD; Regina M. Carney, MD; Goldie S. Byrd, PhD; Jonathan L. Haines, PhD; Gerald D. Schellenberg, PhD; Margaret A. Pericak-Vance, PhD; Rosalyn Lang, PhD and Alzheimer Disease Genetics Consortium. Poster presentation.

American Society for Human Genetics (ASHG), Baltimore, MD, October 6-10, 2015:
Brian W. Kunkle, Badri Vardarajan, Patrice L. Whitehead, Sophie Rolati, Eden R. Martin, John R. Gilbert, Richard P. Mayeux, Jonathan L. Haines, Margaret A. Pericak-Vance, Gary W. Beecham. Whole-exome sequencing identifies novel candidate genes for early- onset Alzheimer disease. Poster presentation.

Cuccaro ML, Carney RM, Kunkle BW, Vance JM, Whitehead PL, Gilbert JR, Vardarajan BN, Haines JL, Mayeux R, Pericak-Vance MA. SORL1 mutations and Parkinsonian features in early onset Alzheimer's disease families. Poster presentation.

International Congress of Human Genetics, Kyoto, Japan, April 3-7, 2016:
Gary W. Beecham, PhD; Brian W. Kunkle, PhD, MPH; Badri Vardarajan, PhD; Patrice L. Whitehead, BS; Sophie Rolati, MS; Eden R. Martin, PhD; John R. Gilbert, PhD. Novel candidate genes for early-onset Alzheimer disease identified using whole-exome sequencing. Platform Presentation

Pericak-Vance MA, Kunkle BW, Carney RM, Kohli MA, Naj AC, Hamilton KL, Whitehead PL, Cuccaro ML, Vance JM, Byrd G, Beecham GW, Gilbert JR, Haines JL, Martin ER. Targeted sequencing of Late-Onset Alzheimer Disease Loci Identifies Genomic Regions with Potential Functional Variants. Poster presentation.

Alzheimer's Association International Conference (AAIC), Toronto, Canada, July 24-28, 2016:
Derek M. Dykxhoorn, Holly N. Cukier, Brian W. Kunkle, Badri N. Vardarajan, Sophie Rolati, Kara L. Hamilton-Nelson, Martin A. Kohli, Patrice L. Whitehead, Derek J. Van Booven, Rosalyn Lang, Lindsay A. Farrer, Michael L. Cuccaro, Jeffery M. Vance, John R. Gilbert, Gary W. Beecham, Eden R. Martin, Regina M. Carney, Richard P. Mayeux, Gerald D. Schellenberg, Goldie S. Byrd, Jonathan L. Haines, Margaret A. Pericak-Vance, and the Alzheimer Disease Genetics Consortium. *ABCA7* Frameshift Deletion Associated with Alzheimer's Disease in African Americans. Platform Presentation.

Brian W. Kunkle, PhD, MPH[1], Badri N. Vardarajan, PhD[2], Adam C. Naj, PhD[3], Holly N. Cukier, PhD[1], Derek M Dykxhoorn, Ph.D.[1], Sophie Rolati, MS[1], Patrice L. Whitehead, BS[1], Regina M. Carney, MD[1], Michael L. Cuccaro, PhD[1], Jeffery M. Vance, MD, PhD[1], Alzheimer's Disease Genetics Consortium[4], Lindsay A. Farrer, PhD[5], Jonathan L. Haines, PhD[6], Gerard D Schellenberg, PhD[3], Eden R. Martin, PhD[1], Christiane Reitz, MD PhD[2], Gary W. Beecham, PhD[1], Richard Mayeux, MD, MSc[2]and Margaret A. Pericak-Vance, PhD[1] Identification of Novel Candidate Genes for Early-Onset Alzheimer Disease through Integrated Whole-Exome Sequencing and Exome Chip Array Association Analysis. Poster Presentation.

Cuccaro ML, Carney RM, Kunkle BW, Cukier HN, Vardarajan BN, Whitehead PL, Mayeux R, St. George-Hyslop P, Pericak-Vance MA. Pathogenic *SORL1* Mutations and Parkinsonian Features in Alzheimer's disease. Poster Presentation.

American Society for Human Genetics (ASHG), Vancouver, Canada, October 18-22, 2016:
Cukier HN, Gross SP, Kunkle BW, Rolati S, Hamilton-Nelson KL, Dombroski BA, Vardarajan BN, Whitehead PL, Lang R, Martin ER, Beecham GW, Farrer LA, Cuccaro ML, Vance JM, Mayeux RP, Gilbert JR, Carney RM, Byrd GS, Haines JL, Schellenberg GD, Pericak-Vance MA, Dykxhoorn DM. Patient-derived iPSC model of an *ABCA7*

deletion associated with Alzheimer disease. Platform Presentation. (Appendix IV)

Cuccaro ML, Carney RM, Zhang Y, Bohm C, Kunkle BW, Vardarajan BN, Whitehead PL, Cukier HN, Mayeux R, St. George-Hyslop P, Pericak-Vance MA. *Pathogenic SORL1* mutations in Alzheimer's Disease. Poster Presentation.

Alzheimer's Association International Conference (AAIC), London, United Kingdom, July 16-20, 2017:
Cukier HN, Mehta N, Ramirez J, Rolati S, Whitehead PL, Adams LD, Celis K, Carney RM, Vance JM, Cuccaro ML, Byrd GS, Pericak-Vance MA, Dykxhoorn DM. Patient-Derived iPSC Model of an *ABCA7* Frameshift Deletion Associated with Alzheimer's Disease in African Americans. poster presentation.

Christiane Reitz, Min Tang, Dolly Reyes-Dumeyer, Brian Kunkle, Kara Hamilton, Rafael Lantigua, Martin Medrano, Eden Martin, Ivonne Jimenez-Velazquez , Margaret Periczak-Vance, Richard Mayeux, Gary W. Beecham. Collection of Multiplex Families with Unexplained Early-Onset Alzheimer's Disease for Genomic Research. poster presentation.

Katrina Celis, Briseida Feliciano, Larry Adams, Parker Bussies, Carolina Sierra, Kara Hamilton, Farid Rajabli, Angel Chinea, Heriberto Acosta, Jacob McCauley, Jeffery Vance, Michael Cuccaro, Gary Beecham and Margaret Pericak-Vance. The Puerto Rican Alzheimer Disease Initiative (PRADI) Initial Clinical Findings. poster presentation.

Briseida Feliciano, Katrina Celis, Larry Adams, Kara Hamilton, Parker Bussies, Carolina Sierra, Nereida Feliciano, Angel Chinea, Jacob L. McCauley, Heriberto Acosta, Jeffery M. Vance, Michael L. Cuccaro, Gary Beecham, Margaret A. Pericak-Vance. The Puerto Rican Alzheimer Disease Initiative (PRADI): A Multisource Ascertainment Approach. poster presentation.

Farid Rajabli**,** Kara L. Hamilton-Nelson, Briseida E. Feliciano-Astacio, Katrina Celis, Larry Deon Adams, Parker Bussies, Carolina Sierra, Angel Chinea, Heriberto Acosta, Jacob L. McCauley, Jeffery M. Vance, Michael L. Cuccaro, Jonathan L. Haines, William S. Bush, Gary W. Beecham and Margaret Pericak-Vance. The Relevance of ApoE to Alzheimer's Disease in the Presence of Local Ancestry Differences. poster presentation.

International Conference on Alzheimer's & Parkinson's Diseases (AD/PD), Vienna, Austria, Mar 29-Apr 2, 2017:
Cukier HN, Gross SP, Kunkle BW, Rolati S, Hamilton-Nelson KL, Whitehead PL, Martin ER, Beecham GW, Cuccaro ML, Vance JM, Mayeux RP, Gilbert JR, Byrd GS, Haines JL, Schellenberg GD, Pericak-Vance MA, Dykxhoorn DM. Patient-derived iPSC model of an *ABCA7* deletion associated with Alzheimer disease in African Americans. Poster Presentation.

American Society for Human Genetics (ASHG), Orlando, Florida, October 17-21, 2017:
Cukier HN, Johnson FS, Garcia Serje C, Carney RM, Vance JM, Cuccaro ML, Pericak-Vance MA, Dykxhoorn DM. A patient-derived iPSC model of a rare *TTC3* mutation segregating with Alzheimer's disease.

F. Rajabli, B.E. Feliciano-Astacio, K. Celis, K.L. Hamilton-Nelson, L.D. Adams, A.R. Betancourt, H. Acosta, A. Chinea, G.S Bird, C. Reitz, R. Mayeux, J.M. Vance, M.L. Cuccaro, J.L. Haines, M.A. Pericak-Vance, G.W. Beecham. African haplotypic background mitigates the effect of APOE ε4 risk allele in Alzheimer disease. Poster presentation

Alzheimer's Association International Conference (AAIC), Chicago, Illinois, July 22-26, 2018:
Holly N. Cukier, Francelethia S. Johnson, Juliana Ramirez, Patrice L. Whitehead, Larry D. Adams, Regina M. Carney, Jeffery M. Vance, Michael L. Cuccaro, Margaret A. Pericak-Vance, Derek M. Dykxhoorn. A Stem Cell Model of a Rare, Segregating *TTC3* Mutation. Poster presentation

Gary W. Beecham, Badri Vardarajan, Elizabeth Blue, William Bush, James Jaworski, Sandra Barral, Brian Kunkle, Eden Martin, Christiane Reitz, Cornelia van Duijn, Jonathan L. Haines, Gerard D. Schellenberg, Richard P. Mayeux, Ellen Wijsman, and Margaret A. Pericak-Vance, for the Alzheimer's Disease Sequencing Project. Whole-genome sequencing in non-Hispanic white families implicates rare variation in late-onset Alzheimer disease risk. Poster presentation

CONCLUSION:

Mutations in *APP*, *PSEN1* and *PSEN2* lead to familial EOAD and accounting for 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for this form of Alzheimer disease unexplained. We performed Whole-Exome Sequencing (WES) on 55 individuals in 19 Caribbean Hispanic EOAD families and 51 Non-Hispanic White EOAD cases previously screened negative for *APP*, *PSEN1* and *PSEN2* to search for rare variants contributing to risk for EOAD. Variants were filtered for segregating, conserved and functional rare variants (MAF<0.1%) assuming both autosomal and X-linked dominant models. 125 rare, segregating, conserved and functional variants passed our stringent filtering criteria for selection of follow-up genotyping candidates. These variants have undergone follow-up genotyping for segregation in the families and for presence in a cohort of 500 Hispanic cases and controls.

20 top candidate variants were identified from this follow-up genotyping, including a 44 base-pair deletion in the known LOAD gene ABCA7 that was associated with risk of AD in several follow-up cohorts. They include 8 variants that show perfect segregation with AD status in the families and are absent in population controls. These variants are in the genes *MYO3A, AAAS, DICER1, YIPF1, ACAP1, LLGL2, BPIFB2*, and *ABCG2*. An additional 11 variants were identified as follow-up candidates based on them showing near complete segregation (absent in one or a few familial cases) and being absent in all familial and sporadic controls. These variants are in the genes *GPR26*, *ERCC*6, *OR5M9, DNAH3, MYOCD, KIF17, TICRR, PLXNB2, LAMA2, SNRNP48, and GLB1L2.* Follow-up of these 20 top candidates in large cohorts of Hispanic, AA and NHW AD cases and controls helped prioritize several top candidate genes and confirmed the association of the deletion in ABCA7 to increased risk of AD. We also identified several rare coding variants in *SORL1*, *PSEN1*, and *MAPT* in EOAD cases and are investigating a potential link between *SORL1* and Parkinsonism in *SORL1* carriers. Finally, we Identified 5 additional candidate EOAD genes (*HSPG2, DOCK3, OGT, CLSTN1, and PARK2*) through identification of NHW EOAD cases with shared rare coding variants with damaging potential in genes interacting with known EOAD genes. We have identified and published or submitted manuscripts on variants identified in these analyses. We continue to identify candidate risk genes for EOAD, including an endocytic gene, *PSD2*, which we find to be significant in an analysis of EOAD exome chip association data from the ADGC. A comparison of these results to our EOAD WES sequencing identified two NHW cases with rare, damaging, missense variants in the *PSD2* gene.

We also are following up our most promising results in iPSC analysis and have ascertained patient samples from the Hussman Institute Human Genomics (HIHG) cohort and the Columbia University cohort with genetic variants in SORL1, SEC16A, and ABCA7 for these analyses. PBMCs have been reprogrammed into iPSC from the patient samples and validated for pluripotency and karyotype. In addition, we have optimized how to measure the secreted amyloid beta 40 and 42 products, internal tau, measure apoptosis rate with the LIVE/DEAD assay and quantified how the young neurons grow and mature with the incucyte zoom machine. Initial results demonstrate that the ABCA7 lines may produce more toxic forms of amyloid beta and have a slower rate of neuronal maturity, as measured by shorter neurite length in day 35-40 neurons compared to controls. Furthermore, preliminary data shows that both the *ABCA7* and *SORL1* alterations may make the cells more vulnerable to cell death, as shown through the LIVE/DEAD assay.

ABCA7 frameshift deletion associated with Alzheimer disease in African Americans

OPEN

Holly N. Cukier, PhD*
Brian W. Kunkle, PhD, MPH*
Badri N. Vardarajan, PhD*
Sophie Rolati, MS
Kara L. Hamilton-Nelson, MPH
Martin A. Kohli, PhD
Patrice L. Whitehead, BS
Beth A. Dombroski, PhD
Derek Van Booven, BS
Rosalyn Lang, PhD
Derek M. Dykxhoorn, PhD
Lindsay A. Farrer, PhD
Michael L. Cuccaro, PhD
Jeffery M. Vance, MD, PhD
John R. Gilbert, PhD
Gary W. Beecham, PhD
Eden R. Martin, PhD
Regina M. Carney, MD
Richard Mayeux, MD, MSc
Gerard D. Schellenberg, PhD
Goldie S. Byrd, PhD
Jonathan L. Haines, PhD
Margaret A. Pericak-Vance, PhD
For the Alzheimer's Disease Genetics Consortium

Correspondence to Dr. Pericak-Vance: mpericak@med.miami.edu

Supplemental data at Neurology.org/ng

ABSTRACT

Objective: To identify a causative variant(s) that may contribute to Alzheimer disease (AD) in African Americans (AA) in the *ATP-binding cassette, subfamily A* (*ABC1*), *member 7* (*ABCA7*) gene, a known risk factor for late-onset AD.

Methods: Custom capture sequencing was performed on ~150 kb encompassing *ABCA7* in 40 AA cases and 37 AA controls carrying the AA risk allele (rs115550680). Association testing was performed for an *ABCA7* deletion identified in large AA data sets (discovery n = 1,068; replication n = 1,749) and whole exome sequencing of Caribbean Hispanic (CH) AD families.

Results: A 44-base pair deletion (rs142076058) was identified in all 77 risk genotype carriers, which shows that the deletion is in high linkage disequilibrium with the risk allele. The deletion was assessed in a large data set (531 cases and 527 controls) and, after adjustments for age, sex, and *APOE* status, was significantly associated with disease (p = 0.0002, odds ratio [OR] = 2.13 [95% confidence interval (CI): 1.42–3.20]). An independent data set replicated the association (447 cases and 880 controls, p = 0.0117, OR = 1.65 [95% CI: 1.12–2.44]), and joint analysis increased the significance ($p = 1.414 \times 10^{-5}$, OR = 1.81 [95% CI: 1.38–2.37]). The deletion is common in AA cases (15.2%) and AA controls (9.74%), but in only 0.12% of our non-Hispanic white cohort. Whole exome sequencing of multiplex, CH families identified the deletion cosegregating with disease in a large sibship. The deleted allele produces a stable, detectable RNA strand and is predicted to result in a frameshift mutation (p.Arg578Alafs) that could interfere with protein function.

Conclusions: This common *ABCA7* deletion could represent an ethnic-specific pathogenic alteration in AD. ***Neurol Genet* 2016;2:e79; doi: 10.1212/NXG.0000000000000079**

GLOSSARY

AA = African Americans; **ABC** = ATP-binding cassette; **AD** = Alzheimer disease; **CH** = Caribbean Hispanic; **CI** = confidence interval; **GATK** = Genome Analysis Toolkit; **GWAS** = genome-wide association study; **OR** = odds ratio; **SNV** = single-nucleotide variant.

Alzheimer disease (AD) is the leading cause of dementia in the elderly. AD occurs at a higher frequency in minority populations, with estimates of AD being twice as frequent in African Americans (AA) compared with non-Hispanic white (NHW) populations.[1,2] *APOE* was the first gene associated with AD and the *ε4* allele confers an increased risk across populations.[3,4] Although *APOE ε4* occurs more frequently in AA than NHW, paradoxically, it has a lower effect

*These authors contributed equally to the manuscript.

From the John P. Hussman Institute for Human Genomics (H.N.C., B.W.K., S.R., K.L.H.-N., M.A.K., P.L.W., D.V.B., D.M.D., M.L.C., J.M.V., J.R.G., G.W.B., E.R.M., R.M.C., M.A.P.-V.), Department of Neurology (H.N.C., J.M.V., M.A.P.-V.), Dr. John T. Macdonald Foundation Department of Human Genetics (D.M.D., M.L.C., J.M.V., J.R.G., G.W.B., E.R.M., R.M.C.), Miller School of Medicine, University of Miami, FL; The Taub Institute for Research on Alzheimer's Disease and the Aging Brain (B.N.V., R.M.), Gertrude H. Sergievsky Center, Departments of Neurology, Psychiatry, and Epidemiology, College of Physicians and Surgeons, Columbia University, New York, NY; Department of Pathology and Laboratory Medicine (B.A.D., G.D.S.), University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; Department of Biology (R.L., G. S.B., M.A.P.-V.), North Carolina A&T State University, Greensboro, NC; Departments of Medicine, Neurology, Ophthalmology, Genetics & Genomics, Epidemiology, and Biostatistics (L.A.F.), Boston University, MA; and Department of Epidemiology and Biostatistics (J.L.H.), Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, OH.

Coinvestigators are listed at Neurology.org/ng.

Funding information and disclosures are provided at the end of the article. Go to Neurology.org/ng for full disclosure forms. The Article Processing Charge was paid by the authors.

size in AA.[5–7] Therefore, while differing ethnicities share risk genes and alleles, the consequences may be different in distinct populations.

Recent studies have identified >20 additional loci associated with late-onset AD, including *ABCA7*.[8–10] Although *ABCA7* was first implicated in NHW, a genome-wide significant signal was also detected in AA individuals at rs115550680, a position in linkage disequilibrium with the NHW genome-wide association study (GWAS) hits.[8,9,11] The AA allele confers a higher risk ($p = 2.21 \times 10^{-9}$, odds ratio [OR] = 1.79 [95% confidence interval (CI): 1.47–2.12]) than the most significantly associated alleles in NHW.[7–9,11] The effect size of the AA *ABCA7* allele is comparable to *APOE ε4* in AA ($p = 5.5 \times 10^{-47}$, OR = 2.31 [95% CI: 2.19–2.42]).[7,11] To date, there is no evidence of a functional consequence of the AA *ABCA7* risk allele.[12–15] Therefore, targeted sequencing of *ABCA7* was performed to identify potential causative variants. A frameshift deletion was found associated with AD in AA, but was virtually absent in NHW. Thus, this deletion potentially represents a common, ethnic-specific, and likely pathogenic alteration that confers risk to AD.

METHODS

Standard protocol approvals, registrations, and patient consents. All the individuals ascertained for this study provided written informed consent prior to their inclusion. If a study participant was not competent to provide consent, the immediate next of kin or a legal representative provided written consent on the behalf of the participant. All participants were ascertained using a protocol that was approved by the appropriate Institutional Review Board. Oversight of this study falls under the University of Miami Institutional Review Board #20070307.

Sample collection. ***African Americans.*** Individuals were ascertained for this study after they provided informed consent at the John P. Hussman Institute for Human Genomics (HIHG) at the University of Miami Miller School of Medicine (Miami, FL), North Carolina A&T State University (Greensboro, NC), and Case Western Reserve University (Cleveland, OH) for the HIHG data set. Each of the participants was ascertained using the protocol approved by the proper institutional review boards (IRBs). Patients were collected for this study over the course of 10 years, with IRB protocols and amendments being approved at each stage. For the HIHG cohort (discovery), 539 cases were ascertained (415 women and 124 men, mean age at onset 74.0 years [SD 8.5]) and 529 controls (403 women and 126 men, mean age at examination 73.1 years [SD 5.4]). The complete HIHG case-control AA cohort (n = 1,068) included 47 relatives, giving 1,021 independent (unrelated) individuals available for analysis.

Samples from the Alzheimer's Disease Genetics Consortium (ADGC) were collected as described previously.[11] For the ADGC cohort (replication), 687 unrelated cases were ascertained (499 women and 188 men, mean age at onset 78.7 years [SD 8.5]) and 1,062 unrelated controls (774 women and 288 men, mean age at examination 78.6 years [SD 6.7]). This subset of the ADGC cohort was independent from the HIHG cohort.

For both HIHG and ADGC data sets, participants underwent rigorous phenotyping and diagnostic criteria following those of the National Institute of Neurological and Communicative Disorders and Stroke—Alzheimer's Disease and Related Disorders Association, as described previously.[11,15,16] Furthermore, the cognitive status of controls was measured with either the Mini-Mental State Examination (MMSE)[17] or the Modified Mini-Mental State (3 MS)[18] and the Clinical Dementia Rating Scale, which assesses functional decline.[19] All the individuals in both cohorts enrolled self-identified as African American. These data were confirmed by analysis of existing GWAS data.[11]

Caribbean Hispanics. Nineteen multiplex Caribbean Hispanic (CH) families initially recruited as a part of the Genetic Epidemiology of Alzheimer's Disease In Hispanics family study at Columbia University were included. A total of 49 cases and 8 unaffected relatives were involved in this study. Information about patient recruitment, demographics, and clinical phenotyping has been published previously.[20] Each family has at least one member with early-onset AD (age at onset <65 years old).

Custom capture and whole exome sequencing. Custom sequence capture was performed on 77 HIHG samples of African American ancestry (40 cases and 37 controls) all with the AA risk allele. Probes were selected using the Agilent SureDesign program across the region (chr19:983277-1133190, ~150 kb; Agilent Technologies, Santa Clara, CA). Fourteen thousand six hundred thirty-six probes were chosen at a 3× density with the moderately stringent masking setting to cover 84.8% of the region. For whole exome sequencing, samples from the 19 CH families were used (46 cases and 6 unaffected relatives). Capture and sequence library construction was performed on a Sciclone G3 NGS Workstation (Caliper Life Sciences, PerkinElmer, Waltham, MA) using the SureSelect Human All Exon 50 Mb Kit (Agilent Technologies) and the Paired-End Multiplexed Sequencing library kit (Illumina, San Diego, CA) for sequence library preparation. All samples were run on the Illumina HiSeq 2000 and paired-end 2 × 100 sequencing was performed. The sequencing data were processed using the Illumina Real-Time Analysis base calling pipeline version 1.8. The Burrows-Wheeler Aligner was used to map sequences to the hg19 human reference genome, and variant calling was performed with the Genome Analysis Toolkit (GATK[21,22]). GATK parameters included base quality score recalibration and duplicate removal.[23] The data were evaluated for deletions and insertions by alignment with Bowtie2 and analysis using the Pindel program.[24,25]

Sanger sequencing. Both the *ABCA7* deletion (rs142076058) and the AA *ABCA7* risk allele (rs115550680) were sequenced using traditional Sanger sequencing. Custom primers were designed with the Primer3 v4.0 program (http://fokker.wi.mit.edu/primer3/input.htm). For the deletion, primers were selected to flank the 44-base pair (bp) deletion to perform Sanger sequencing for validation (deletion-F: AAATCTTC CCGCCTTGAGAT, deletion-R: GGAGCTTAGGGTGC AGCTC). PCR experiments were set up with 1.5 mM $MgCl_2$, 1.6 M betaine, and touchdown PCR was performed. PCR experiments resulted in amplicons of either 450 or 406 bp. Sequencing of the AA risk allele was performed with the

following primers (rs115550680-F: GCCAATATGGCAAAACCATC, rs115550680-R: TCCAAAACCCTGTGATAGCC) to generate a 245-bp amplicon. PCR reactions were set up with 2 mM $MgCl_2$ and touchdown PCR was performed. Sequencing reactions were performed using the Big Dye Terminator v3.1 (Life Technologies, Carlsbad, CA), reactions were run on a 3730xl DNA Analyzer (Life Technologies), and results were evaluated using the Sequencher v4.10.1 program (Gene Codes Corporation, Ann Arbor, MI).

TaqMan SNP genotyping analysis. Both the *ABCA7* deletion (rs142076058) and the AA *ABCA7* risk allele (rs115550680) were evaluated using the TaqMan single-nucleotide polymorphism (SNP) Genotyping Assays (Life Technologies). The *ABCA7* deletion was evaluated by a custom-designed TaqMan SNP Genotyping Assays designed to recognize the presence or absence of the deletion. This assay had to be ordered as a "non-Human Assay" (forward primer: GCCTGGATCTACTCCGTGAC, reverse primer: GAGGCAGCTGAGGAACCA, FAM probe: GAGACGCGGCTGG—identifies when the sequence is deleted, VIC probe: CGCCATGGGGCT—wild-type allele). Samples were amplified for 40 cycles and, when amplification was low, an additional 20 cycles was added. The plates were read on the 7900HT Fast Real-Time PCR machine (Applied Biosystems, Foster City, CA), and data were analyzed with the SDS v2.4 software.

RNA isolation and real-time PCR. RNA was isolated from blood collected in PAXgene blood RNA tubes (PreAnalytiX, Hombrechtikon, Switzerland) from 10 samples with and 10 samples without the *ABCA7* deletion and extracted following the manufacturer's standard protocol. RNA was quantified on the 2100 BioAnalyzer (Agilent Technologies) and was required to have an RIN ≥6. Complementary DNA (cDNA) libraries were generated using the iScript Reverse Transcriptase Supermix for RT-qPCR kit (BIO-RAD). PCR primers were designed to amplify the cDNA across the deletion (cDNA-F: TGTTCCTGCGTGTGCTGA, cDNA-R: AGCAGGAAGCTCTGGGTCAC) and the resulting PCR products resolved on a 2% agarose gel. The wild-type allele results in an amplicon of 316 bp, whereas the allele with the deletion produces a 272-bp amplicon.

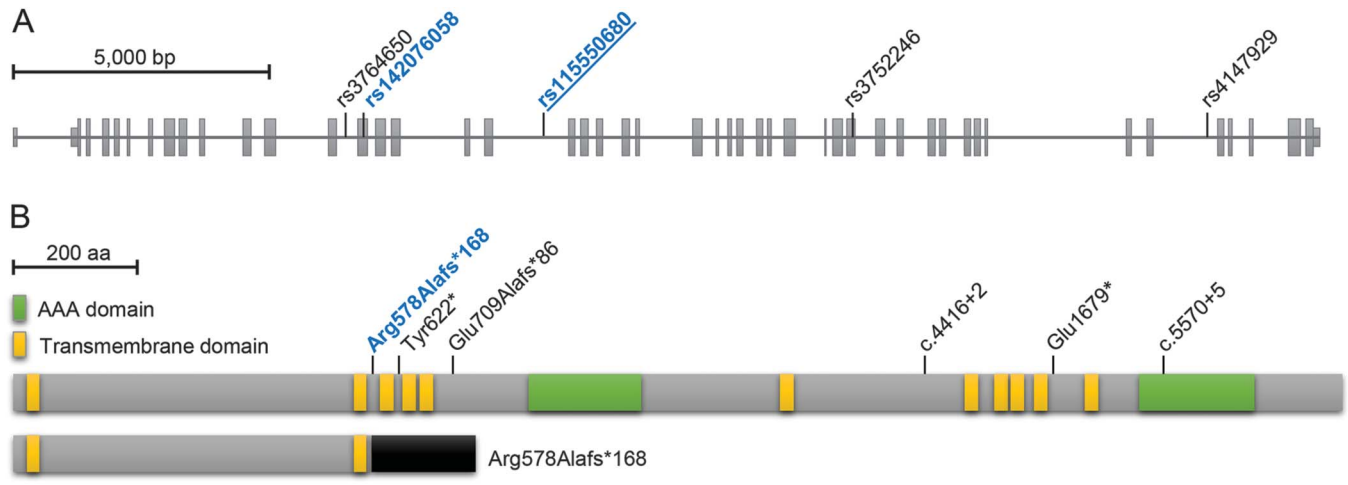
Statistical analysis. The GENMOD program, as part of the SAS/STAT software, was used to perform the association tests under a logistic regression model. Association tests were performed with adjustments for age, sex, *APOE* status, and relatedness between samples (SAS Institute, Cary, NC). Conditional analysis was performed in PLINK.[26] The Fisher exact test was used to evaluate the differences in the alleles frequency of the deletion between African and European populations reported in the ExAC database.

RESULTS

We selected 40 AA AD cases and 37 AA controls (aged >65 years) carrying the AA risk allele, rs115550680, to perform custom massively parallel sequencing of a ~150-kb region that includes *ABCA7* and 8 flanking genes and a small nuclear RNA. Samples were sequenced to an average depth of over 1,000× and evaluated for single-nucleotide variants (SNVs) and insertions and deletions. One thousand one hundred twenty SNVs were detected by sequencing with 11 variants showing different frequencies in cases and controls ($p < 0.1$, table e-1 at Neurology.org/ng). In addition, a 44-base pair (bp) deletion (rs142076058, p.Arg578Alafs) located ~3.5 kb upstream of the AA risk allele was identified in all 77 individuals, which suggests that it is in high linkage disequilibrium with the risk allele (figure 1).

To further evaluate the rs142076058 deletion, a custom TaqMan genotyping assay was designed to evaluate the deletion in our larger AA cohort,

Figure 1 Location of the deletion in the *ABCA7* gene and protein



(A) *ABCA7* gene (chr19:1,040,103-1,065,571, hg38), 44-base pair deletion (blue), African American risk allele (blue and underlined[11]), and 3 non-Hispanic white (NHW) risk alleles (rs3764650,[8] rs3752246,[9] and rs4147929[10]). (B) Wild-type ABCA7 protein (2,146 amino acids) and the location of frameshift deletion (blue) identified in this study. Below, the protein predicted to be generated from deletion would contain only 2 of the 11 transmembrane domains (yellow) and neither of the 2 AAA domains (green), but incorporate 168 aberrant amino acids (black). The remaining frameshift, nonsense, and splicing variants designated are rare alterations (<1% minor allele frequency) previously reported in NHW populations to be associated with Alzheimer disease.[33-35]

Table 1 Association testing of the deletion in African American cohorts

	Samples with deletion/ total samples (%)	OR	95% CI	Pr > |Z|
HIHG				
Cases	86/531 (16.2)	2.13	1.42-3.20	0.0002
Controls	49/527 (9.3)			
ADGC				
Cases	63/447 (14.9)	1.65	1.12-2.44	0.0117
Controls	88/880 (10.0)			
Joint analysis				
Cases	149/978 (15.2)	1.81	1.38-2.37	1.414×10^{-5}
Controls	137/1,407 (9.7)			

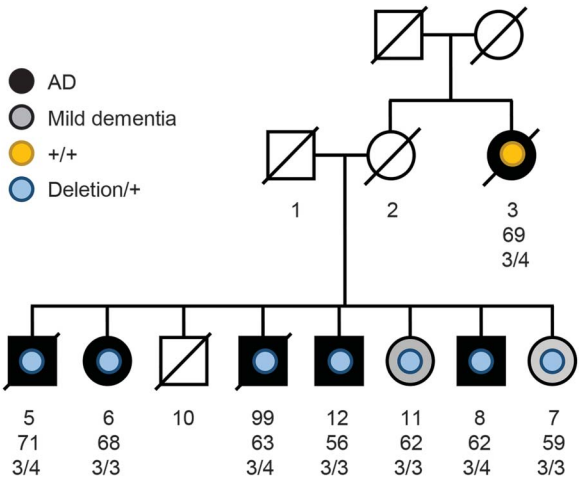
Abbreviations: ADGC = Alzheimer's Disease Genetics Consortium; CI = confidence interval; HIHG = John P. Hussman Institute for Human Genomics, University of Miami.

designated as HIHG. After adjustments for age, sex, and *APOE* status, the deletion was found to be significantly associated with AD ($p = 0.0002$, OR = 2.13 [95% CI: 1.42–3.20], table 1) in 531 cases and 527 controls. The deletion occurred in 9.3% of control individuals but in 16.2% of AD cases. A subset of 8 individuals were pathologically confirmed cases of AD; 2 were found to carry the deletion whereas the remaining 6 did not. The AA risk allele was also genotyped in this data set. The risk allele was significantly associated with AD ($p = 0.0005$, OR = 2.07 [95% CI: 1.38–3.13]) and in linkage disequilibrium with the deletion ($D' = 1.000$, $r^2 = 0.995$, tables e-2 and e-3). The top SNPs in *ABCA7* reported previously in NHW studies were also found to be in linkage disequilibrium with the deletion (table e-4).[8,9,11] Genotyping was also performed on our NHW AD samples (n = 3,275), but only 4 individuals were identified with the deletion (0.12%), all of whom carried the AA risk allele, which indicates that the genomic fragment carrying the deletion in these individuals may be of African descent.

An independent AA data set from the Alzheimer's Disease Genetics Consortium (ADGC) was evaluated in the same manner, and the deletion was again associated with AD ($p = 0.0117$, OR = 1.65 [95% CI: 1.12–2.44], table 1), occurring in 10.0% of control individuals and 14.9% of AD cases. Joint analysis of the 2 cohorts increased the strength of the association ($p = 1.414 \times 10^{-5}$, OR = 1.81 [95% CI: 1.38–2.37], table 1). Association testing was also performed for each data set without *APOE* adjustment; more significant results were obtained with *APOE* adjustment, demonstrating that *APOE* did not influence the association (table e-5). Examination of the ages of cases and controls with and without the deletion did not find a significant difference between any of these groups (table e-6).

To examine the association with AD in another ethnic group with a high level of African ancestry (~42%), we evaluated whole exome sequencing data on 19 CH families from the Dominican Republic with multiple affected AD participants.[27] In addition to a relatively high level of African ancestry, CHs are highly inbred and have a high incidence of AD, and are thus enriched for AD genetic risk factors.[28] We independently identified the same 44-bp deletion from whole exome sequencing of 3 affected individuals from a large CH family. Subsequent examination of the family revealed that the deletion segregated in a large sibship in the family (figure 2). Both the deletion and AA risk allele were isolated in all 7 siblings who clinically presented in a range from AD (individuals 5, 6, 8, 12, and 99) to milder stages of dementia (individuals 7 and 11). Haplotype analysis around the *ABCA7* deletion using SNP data in the 1-Mb flanking region on the family revealed that an affected aunt who does not carry the deletion, individual 3 (figure 2), has distinct ancestral haplotypes from the family members in the large sibship with the deletion (table e-7). This finding suggests that individual 3's AD phenotype can be attributed to other genetic factors and that the *ABCA7* deletion is highly penetrant in the sibship. Because several members of this CH family were known to have early-onset AD (age at onset <65 years), we examined the entire AA AD cohort (both HIHG and ADGC) to determine whether there was an effect of the deletion on age at onset in AD. We found no

Figure 2 Pedigree of an AD family from the Dominican Republic with the *ABCA7* deletion



Family 360 has 6 individuals diagnosed with Alzheimer disease (AD) and 2 individuals presenting with mild dementia. The numbers beneath each individual in the pedigree represent the individual's sample number, the age at onset of AD (for AD cases) or the age at examination, and the *APOE* genotype. All 7 siblings carry the *ABCA7* deletion and the African Americans (AA) risk allele, whereas individual 3 had neither the deletion nor the AA risk allele.

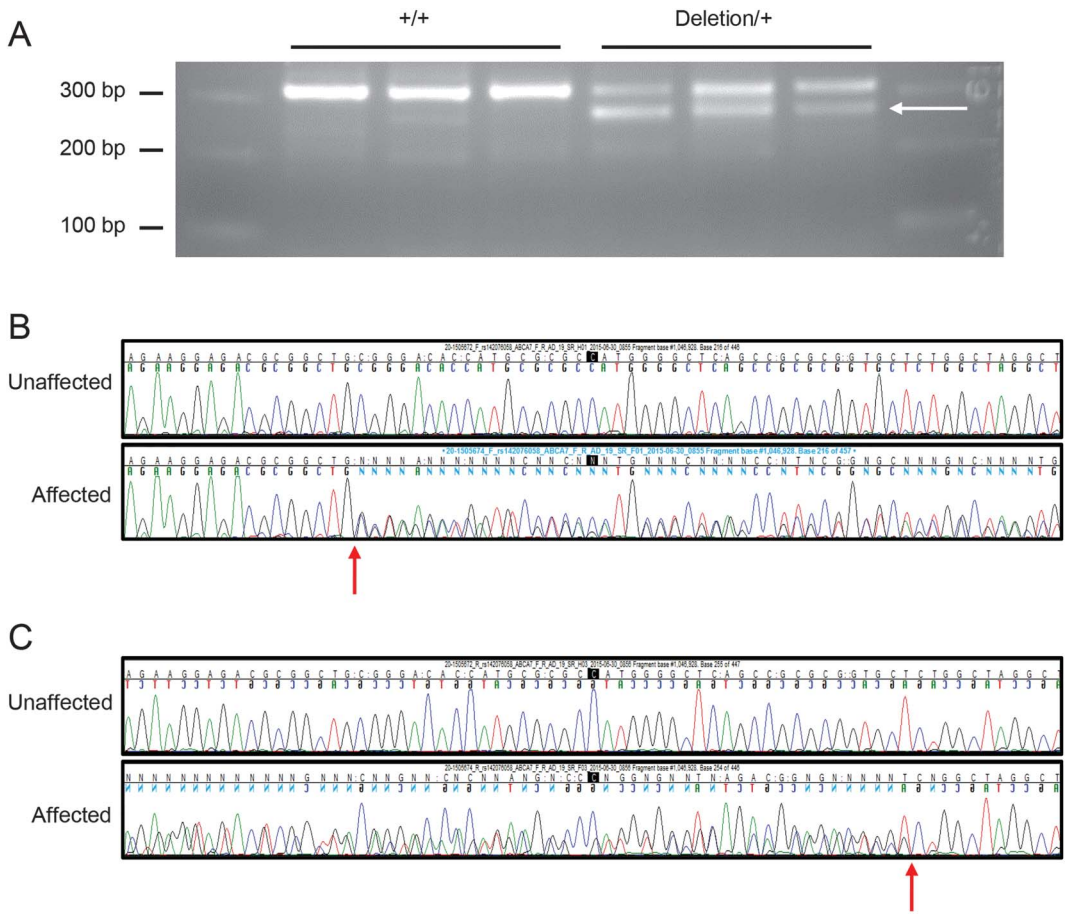
difference in the age at onset in cases with the deletion (75.6 years [SD 9.6]) compared with cases absent for the deletion (76.8 years [SD 8.7], $p = 0.09$).

To determine whether the *ABCA7* allele with the deletion was being transcribed, RNA was isolated from the blood of AA individuals both with and without the *ABCA7* deletion. Reverse transcription PCR across the deletion region demonstrated that the allele carrying the deletion is transcribed and produces a stable detectable RNA strand (figure 3).

This deletion was reported previously in the Exome Aggregation Consortium (ExAC: http://exac.broadinstitute.org) [June 2015], a repository of 60,706 unrelated individuals from 6 distinct ethnic groups. In the ExAC data set, the deletion was found in 7.77% of individuals of African ancestry and 0.95% of Latino individuals, but was absent from individuals of European ancestry. This difference in population frequencies between the African and European populations was highly significant ($p < 1 \times 10^{-10}$).

DISCUSSION We identified a 44-bp deletion in *ABCA7* that is associated with AD in individuals of African ancestry. Although the deletion did occur in unaffected individuals, it was found at a higher frequency in individuals with AD (15.2% of cases vs 9.74% of controls), implicating it as a risk factor for disease. This reaffirms that the deletion is likely to be of African ancestry. Furthermore, the combined cases from the HIHG and ADGC data sets had the deletion at a frequency of 15.2%, approximately twice as high as that identified with through ExAC African populations (7.8%), lending additional evidence of a relationship to disease. The deletion was also independently identified in an AD family from the Dominican Republic, a population that has a relatively high level of African ancestry, 41.8%.[27] Examination of the linkage disequilibrium

Figure 3 **Deletion allele produces an RNA transcript**



(A) Real-time PCR from cDNA of 3 samples without the *ABCA7* deletion (+/+) and 3 samples heterozygous for the *ABCA7* deletion (deletion/+). All samples produce an amplicon of 316 base pairs, but only the samples with deletion generate a lower, 272-base pair amplicon (arrow). (B) Sanger sequencing from the 5′ end of the deletion in an African American control lacking the deletion and Alzheimer disease (AD)-specific line heterozygous for the deletion. The arrow denotes where the deleted allele begins to be out of frame with the wild-type allele C. Sanger sequencing from the 3′ end of the deletion from the same control and AD individuals.

of the deletion with the top 3 previously reported SNPs found a high D′ across all locations, but only a large r^2 with the African-specific risk allele (table e-4), further supporting that distinct alleles confer AD risk in different ethnicities.[8,9,11]

The deletion is predicted to cause a frameshift at amino acid 578, encoding for 168 incorrect amino acids before stopping prematurely compared with the largest isoform that generates a protein of 2,146 amino acids (figure 1B). Because we were able to detect RNA from the allele with the deletion, it is possible that this RNA generates an aberrant protein that interferes with the wild-type 2,146 amino acid protein. Within the first 578 amino acids, 2 transmembrane regions are conserved and would be maintained by the mutated protein (figure 1B). However, both AAA domains and 9 additional transmembrane domains would be predicted to be lost in this truncated protein, and thereby interfere with the protein's function of exporting the lipid phosphatidylserine.[29] Alternatively, the shortened transcript may be subjected to nonsense-mediated decay (NMD), as was seen in the Glu709fs alteration identified in NHW.[30] Although some loss-of-function variants in *ABCA7* were identified in NHW populations that may contribute to AD pathogenicity, these are rare variants and may only partially contribute to the NHW GWAS signal.[30–32] A few previously reported loss-of-function variants have demonstrated a functional consequence including the Glu709fs variant undergoing NMD, and the c.5570+5G>C alteration led to aberrant splicing[30,31] (figure 1B). Therefore, this study may be the first to connect a potentially pathogenic and common alteration with a GWAS signal in *ABCA7*.

ABCA7 is a member of the ATP-binding cassette (ABC) transporter family, a large group of 49 genes that encode for membrane proteins that facilitate the movement of substrates across cell membranes.[33,34] *ABCA7* is expressed in the brain in neurons and microglia.[35,36] There is evidence both in patients and animal models demonstrating that inadequate levels of *ABCA7* may be directly correlated with Alzheimer pathogenesis.[37–39] The ABCA7 protein is involved in the processing of amyloid precursor protein.[40] In addition, evidence has shown that ABCA7 acts in the phagocytic pathway through extracellular signal-regulated kinase signaling.[e1,e2] *ABCA7* is not the only ABC transporter gene linked to AD; *ABCA1*, *ABCB1*, *ABCC1*, *ABCG1*, *ABCG2*, and *ABCG4* are all implicated in Aβ regulation.[e3–e8] Furthermore, a study identified rare loss-of-function alterations in NHW patients diagnosed with Parkinson disease, including specific variants previously reported in AD individuals, which demonstrates that this gene may contribute to the risk of multiple neurodegenerative disorders.[e9]

Therefore, the results of this study demonstrate that there is a 44-bp deletion in *ABCA7* that is associated with AD and in linkage disequilibrium with the previously identified AA risk allele. The deletion was relatively frequent in our large AA AD cohorts, independently identified in 1 of 19 CH AD families, and virtually absent from our large NHW AD cohort. Thus, the deletion could represent a common, ethnic-specific alteration that confers risk of AD in populations with African ancestry.

AUTHOR CONTRIBUTIONS

H.N.C., B.W.K., B.N.V., M.A.K., J.R.G., R.M., and M.A.P.-V. conceived and designed the experiments. R.L., M.L.C., J.M.V., R.M.C., G.S.B., J.L.H., and M.A.P.-V. acquired and assessed the HIHG samples. L.A.F., R.M., and G.D.S. acquired and assessed the ADGC samples. S.R., P.L.W., and B.A.D. performed custom capture sequencing, exome sequencing, Sanger sequencing, and TaqMan genotyping. H.N.C., B.W.K., B.N.V., S.R., K.L.H.-N., B.A.D., D.V.B., M.A.K., G.W.B., J.R.G., and M.A.P.-V. analyzed the sequencing and TaqMan data. B.W.K., K.L.H.-N., and M.A.P.-V. performed the statistical analysis. H.N.C., S.R., and D.M.D. performed RT-PCR. H.N.C. drafted the manuscript. B.W.K., B.N.V., D.M.D., J.M.V., G.W.B., M.A.K., R.M., G.D.S., J.L.H., and M.A.P.-V. edited the manuscript. The authors jointly discussed the experimental results throughout the duration of the study. All authors read and approved the final manuscript.

ACKNOWLEDGMENT

The authors acknowledge Doris Caldwell and Larry Deon Adams for their assistance in patient ascertainment and data management, and Anthony Griswold for his help in bioinformatics.

STUDY FUNDING

This research was supported by grants from the NIH [R01 AG027944, R01 AG028786 to M.A.P.-V., R01 AG019085 to J.L.H., P20 MD000546, R01 AG28786-01A1 to G.S.B., U01-AG032984, RC2-AG036528 to G.D.S., U24 AG026395, U24 AG026390, R01AG037212, R37 AG015473, R01AG015473 to R.M., R01 AG009029 to L.A.F., U01-AG016976 to W.A.K., U24-AG021886 to T.M.F., R01AG009956, RC2 AG036650 to K.H., UO1 AG06781, UO1 HG004610 to E.L., 5R01AG20688 to M.D.F., P50 AG005133, AG030653 to I.K., R01 AG1101, R01 AG030146, RC2 AG036650 to D.A.E., P30AG10161, R01AG15819, R01AG30146, R01AG17917, R01AG15819 to D.A.B., R01AG028786 to J.J.M., R01AG22018, P30AG10161 to L.L.B., P50AG16574, R01 032990, KL2 RR024151 to N.E.T. and N.R.G.R., AG005138 to J.D.B., P50 AG05681, P01 AG03991, P01 AG026276 to A.M.G., and U24-AG041689 to L.S.W., and U19 AG047133 and UF1AG047133]; the Department of Defense [W81XWH-12-1-0013 to M.A.P.-V.]; a joint grant from the Alzheimer's Association [SG-14-312644] and the Fidelity Biosciences Research Initiative to M.A.P.-V.; and the BrightFocus Foundation [A2011048 to M.A.P.-V.].

DISCLOSURE

Dr. Cukier has received research support from the NARSAD Young Investigator Grant (The Brain & Behavior Research Foundation), the Hussman Foundation, and the Alzheimer's Association. Dr. Kunkle reports no disclosures. Dr. Vardarajan has been a consultant for Immuneering Corporation. Ms. Rolati reports no disclosures. Ms. Hamilton-Nelson has received research support from NIH. Dr. Kohli, Ms. Whitehead, Dr. Dombroski, and Mr. Van Booven report no disclosures. Dr. Lang has served on the editorial board of the *Journal of Parkinson's Disease and Alzheimer's Disease*. Dr. Dykxhoorn has received research support from NIAID/NIH. Dr. Farrer has served on the editorial boards of the

American Journal of Alzheimer's Disease & Other Dementias and *Clinical Genetics*; holds a patent for Use of PLXNA4 as a drug target and biomarker for Alzheimer disease; has been a consultant for Novartis Pharmaceuticals, Gerson Lerman, and Guidepoint Global; has received research support from NIH, the Fidelity Foundation, and the Thome Memorial Foundation; and has been a consultant for Finnegan & Associates LLP regarding legal proceedings. Dr. Cuccaro has served on the editorial board of *Child Psychiatry & Human Development*. Dr. Vance has received honoraria from the University of Alaska (served on an NIH grant's internal review), and NEPTR, Department of Defense; has served on the editorial boards of the *American Journal of Neurodegenerative Diseases* and *Neurology Genetics*; holds patents for Method of detecting Charcot-Marie-Tooth disease type 2A, TRPC6 involved in glomerulonephritis, and Methods for identifying an individual at increased risk of developing coronary artery disease; has received research support from NIH/NINDS and the Hussman Foundation; and has received royalty payments from Duke University. Dr. Gilbert has been a consultant for Glenview NGLG; and has received research support from the Alzheimer's Association. Dr. Beecham has received research support from NIH and the Department of Defense. Dr. Martin has served on the editorial board of *Frontiers in Statistical Genetics and Methodology*; and holds US Patent No. 6697739—Test for Linkage and Association in General Pedigrees: The Pedigree Disequilibrium Test. Dr. Carney reports no disclosures. Dr. Mayeux has received research support through government grants RO1 AG037212 and RF1 AG015473. Dr. Schellenberg has served on the Alzheimer's Association Medical Advisory Board, Society of Progressive Supranuclear Palsy Scientific Advisory Board, Alzheimer Research Consortium Peebler PSP Research Foundation Medical and Scientific Advisory Committee, the Alzheimer's Association External Scientific Advisory Board, the scientific advisory board of the United Kingdom Parkinson Disease Center, the steering committee of the University College London, the Alzheimer's Disease Sequence Project, Structural Variant Work Group, the advisory committee of the Alzheimer's Disease Sequence Project, the Mayo Clinic Rochester Udall Center External Advisory Committee, the Discovery Assessment Panel of the University of Miami Udall Center, and the scientific advisory board of the Oxford Parkinson's Disease Centre; has received travel funding/speaker honoraria from the Alzheimer's Disease Center, CurePSP, the University of California, the 25th Annual Southern California Alzheimer's Disease Research Conference, the University of California Institute for Memory Impairment and Neurological Disorders, Biomarkers in Neuropsychiatric Disorders (Toronto, Canada), NIH, Novartis, the McKnight Brain Institute, the University of Florida, the Florida Alzheimer's Disease Center, the Lou Ruvo Center for Brain Health, the Accelerated Medicines Program Alzheimer's Disease, the University of Virginia, and Columbia University; has served on the editorial boards of the *Journal of Neural Transmission*, the *American Journal of Alzheimer's Disease and other Dementias*, *Neurodegenerative Diseases*, *Current Alzheimer Research*, and *Pathology and Laboratory Medicine International*; is an employee of the University of Pennsylvania; and has received research support from NIA/NIH, CurePSP, and CBD Solutions. Dr. Byrd reports no disclosures. Dr. Haines has received travel funding/speaker honoraria from Novartis; has served on the editorial boards of *Neurogenetics*, *Current Protocols in Human Genetics*, and *Human Molecular Genetics*; has received royalties from John Wiley & Sons; has received research support from the NIH/NIA Consortium for Alzheimer's Sequence Analysis (CASA) and NIH/NEI; and has received royalty payments from Athena Diagnostics. Dr. Pericak-Vance reports no disclosures. Go to Neurology.org/ng for full disclosure forms.

Received January 5, 2016. Accepted in final form March 31, 2016.

REFERENCES

1. Green RC, Cupples LA, Go R, et al. Risk of dementia among white and African American relatives of patients with Alzheimer disease. JAMA 2002;287:329–336.
2. Chin AL, Negash S, Hamilton R. Diversity and disparity in dementia: the impact of ethnoracial differences in Alzheimer disease. Alzheimer Dis Assoc Disord 2011;25:187–195.
3. Strittmatter WJ, Saunders AM, Schmechel D, et al. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A 1993;90:1977–1981.
4. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–923.
5. Murrell JR, Price B, Lane KA, et al. Association of apolipoprotein E genotype and Alzheimer disease in African Americans. Arch Neurol 2006;63:431–434.
6. Ward A, Crean S, Mercaldi CJ, et al. Prevalence of apolipoprotein E4 genotype and homozygotes (APOE e4/4) among patients diagnosed with Alzheimer's disease: a systematic review and meta-analysis. Neuroepidemiology 2012;38:1–17.
7. Reitz C, Mayeux R. Genetics of Alzheimer's disease in Caribbean Hispanic and African American populations. Biol Psychiatry 2014;75:534–541.
8. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 2011;43:429–435.
9. 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.
10. 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.
11. Reitz C, Jun G, Naj A, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E 4, and the risk of late-onset Alzheimer disease in African Americans. JAMA 2013;309:1483–1492.
12. ENCODE Project Consortium, Dunham I, Kundaje A, et al. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.
13. 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;46:310–315.
14. Shihab HA, Rogers MF, Gough J, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics 2015;31:1536–1543.
15. McKhann G, Drachman D, Folstein M. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of the Department of Health and human Services Task Force on Alzheimer's disease. Neurology 1984;34:939–944.
16. 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.
17. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–198.
18. Teng EL, Chui HC. The modified mini-mental state (3MS) examination. J Clin Psychiatry 1987;48:314–318.

19. Morris JC. The clinical dementia rating (CDR): current version and scoring rules. Neurology 1993;43:2412–2414.
20. 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.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25:1754–1760.
22. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–1303.
23. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–498.
24. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009;25:2865–2871.
25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–359.
26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
27. Bryc K, Velez C, Karafet T, et al. Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proc Natl Acad Sci USA 2010;107(suppl 2):8954–8961.
28. 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;17:639–643.
29. Quazi F, Molday RS. Differential phospholipid substrates and directional transport by ATP-binding cassette proteins ABCA1, ABCA7, and ABCA4 and disease-causing mutants. J Biol Chem 2013;288:34414–34426.
30. Cuyvers E, De Roeck A, Van den Bossche T, et al. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. Lancet Neurol 2015;14:814–822.
31. Steinberg S, Stefansson H, Jonsson T, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat Genet 2015;47:445–447.
32. Vardarajan BN, Ghani M, Kahn A, et al. Rare coding mutations identified by sequencing of Alzheimer's disease GWAS loci. Ann Neurol 2015;78:487–498.
33. Kaminski WE, Orso E, Diederich W, Klucken J, Drobnik W, Schmitz G. Identification of a novel human sterol-sensitive ATP-binding cassette transporter (ABCA7). Biochem Biophys Res Commun 2000;273:532–538.
34. Pahnke J, Frohlich C, Krohn M, Schumacher T, Paarmann K. Impaired mitochondrial energy production and ABC transporter function-A crucial interconnection in dementing proteopathies of the brain. Mech Ageing Dev 2013;134:506–515.
35. Langmann T, Mauerer R, Zahn A, et al. Real-time reverse transcription-PCR expression profiling of the complete human ATP-binding cassette transporter superfamily in various tissues. Clin Chem 2003;49:230–238.
36. Kim WS, Guillemin GJ, Glaros EN, Lim CK, Garner B. Quantitation of ATP-binding cassette subfamily-A transporter gene expression in primary human brain cells. Neuroreport 2006;17:891–896.
37. Logge W, Cheng D, Chesworth R, et al. Role of Abca7 in mouse behaviours relevant to neurodegenerative diseases. PLoS One 2012;7:e45959.
38. Vasquez JB, Fardo DW, Estus S. ABCA7 expression is associated with Alzheimer's disease polymorphism and disease status. Neurosci Lett 2013;556:58–62.
39. Kim WS, Li H, Ruberu K, et al. Deletion of Abca7 increases cerebral amyloid-beta accumulation in the J20 mouse model of Alzheimer's disease. J Neurosci 2013; 33:4387–4394.
40. Chan SL, Kim WS, Kwok JB, et al. ATP-binding cassette transporter A7 regulates processing of amyloid precursor protein in vitro. J Neurochem 2008;106:793–804.

SORL1 mutations in early- and late-onset Alzheimer disease

OPEN

Michael L. Cuccaro, PhD*
Regina M. Carney, MD*
Yalun Zhang, PhD
Christopher Bohm, PhD
Brian W. Kunkle, PhD
Badri N. Vardarajan, PhD
Patrice L. Whitehead, BS
Holly N. Cukier, PhD
Richard Mayeux, MD, MSc
Peter St. George-Hyslop, MD
Margaret A. Pericak-Vance, PhD

Correspondence to Dr. Pericak-Vance: mpericak@med.miami.edu

ABSTRACT

Objective: To characterize the clinical and molecular effect of mutations in the sortilin-related receptor (*SORL1*) gene.

Methods: We performed whole-exome sequencing in early-onset Alzheimer disease (EOAD) and late-onset Alzheimer disease (LOAD) families followed by functional studies of select variants. The phenotypic consequences associated with *SORL1* mutations were characterized based on clinical reviews of medical records. Functional studies were completed to evaluate β-amyloid (Aβ) production and amyloid precursor protein (APP) trafficking associated with *SORL1* mutations.

Results: *SORL1* alterations were present in 2 EOAD families. In one, a SORL1 T588I change was identified in 4 individuals with AD, 2 of whom had parkinsonian features. In the second, an SORL1 T2134 alteration was found in 3 of 4 AD cases, one of whom had postmortem Lewy bodies. Among LOAD cases, 4 individuals with either SORL1 A528T or T947M alterations had parkinsonian features. Functionally, the variants weaken the interaction of the SORL1 protein with full-length APP, altering levels of Aβ and interfering with APP trafficking.

Conclusions: The findings from this study support an important role for *SORL1* mutations in AD pathogenesis by way of altering Aβ levels and interfering with APP trafficking. In addition, the presence of parkinsonian features among select individuals with AD and *SORL1* mutations merits further investigation. ***Neurol Genet* 2016;2:e116; doi: 10.1212/NXG.0000000000000116**

GLOSSARY

AAO = age at onset; **Aβ** = β-amyloid; **AD** = Alzheimer disease; **APP** = amyloid precursor protein; **APPsβ** = APP soluble β-secretase; **APPsw** = Swedish APP mutant; **EOAD** = early-onset Alzheimer disease; **ER** = endoplasmic reticulum; **FL-APP** = full-length APP; **HIHG** = John P. Hussman Institute for Human Genomics; **LOAD** = late-onset Alzheimer disease; **PD** = Parkinson disease; ***SORL1*** = sortilin-related receptor; **WES** = whole-exome sequencing.

Alzheimer disease (AD) is the leading cause of dementia in the elderly.[1] Multiple genes have been implicated in risk for both late-onset Alzheimer disease (LOAD; onset >65 years of age) and early-onset Alzheimer disease (EOAD; onset <65 years of age)[2] including the sortilin-related receptor (*SORL1*) gene. Located on chromosome 11q23.2-q24.2, *SORL1* influences the differential sorting of the amyloid precursor protein (APP) and regulation of β-amyloid (Aβ) production, making it biologically plausible for AD risk.[3–9]

Compelling evidence for the involvement of *SORL1* in AD comes from a large meta-analysis of >30,000 individuals, which confirmed that variants in *SORL1* are associated with AD risk.[10]

Supplemental data at Neurology.org/ng

*These authors contributed equally to this work.

From the John P. Hussman Institute for Human Genomics (M.L.C., R.M.C., B.W.K., P.L.W., H.N.C., M.A.P.-V.), University of Miami Miller School of Medicine, FL; Mental Health & Behavioral Science Service (R.M.C.), Bruce W. Carter VA Medical Center, Miami, FL; The Taub Institute for Research on Alzheimer's Disease and the Aging Brain (B.N.V., R.M.), Gertrude H. Sergievsky Center, Departments of Neurology, Psychiatry, and Epidemiology, College of Physicians and Surgeons, Columbia University, New York, NY; Tanz Centre for Research in Neurodegenerative Diseases and Department of Medicine (Y.Z., C.B., P.S.G.-H.), University of Toronto, Ontario, Canada; Cambridge Institute for Medical Research (P.S.G.-H.), Department of Clinical Neurosciences, University of Cambridge, United Kingdom; and Dr. John T. Macdonald Foundation Department of Human Genetics (M.L.C., M.A.P.-V.), Miller School of Medicine, University of Miami, FL.

Funding information and disclosures are provided at the end of the article. Go to Neurology.org/ng for full disclosure forms. The Article Processing Charge was paid by the authors.

Furthermore, whole-exome sequencing (WES) has identified potentially damaging *SORL1* mutations in patients with both EOAD and LOAD.[11,12] Of note, a WES study of a large EOAD cohort found a greater frequency of predicted damaging missense *SORL1* variants in cases vs controls, with this effect enriched among cases with a positive family history.[13] Clearly, rare coding variants in *SORL1* are tied to risk for EOAD and LOAD. Finally, while *SORL1* mutations have been reported in multiple patients with AD, there has been little investigation of clinical phenotypes beyond dementia and age at onset (AAO) among these individuals.

For this study, we examined well-characterized EOAD families using WES to discover AD risk genes. Our efforts focused on clinical characterization of individuals with *SORL1* alterations and investigation of the functional effect of the identified *SORL1* alterations in a series of gene overexpression experiments.

METHODS

Standard protocol approvals, registrations, and patient consents. All participants ascertained for this study gave written informed consent prior to their inclusion. If an individual was not competent to give consent, the immediate next of kin or a legal representative provided written consent on their behalf. All participants were ascertained using a protocol that was approved by the appropriate Institutional Review Board. Oversight of this study falls under the University of Miami Institutional Review Board #20070307.

Sources of participants. EOAD families were ascertained as part of a larger study on AD genetics whose participants were enrolled under protocols previously described.[14,15] Individuals were ascertained for this study after they provided informed consent at the John P. Hussman Institute for Human Genomics (HIHG) at the University of Miami Miller School of Medicine (Miami, FL). The majority of these families were self-reported non-Hispanic whites (N = 47); the remaining families were self-reported African Americans (N = 3). Clinical data from cognitively impaired individuals, including any that changed affection status, were evaluated by the HIHG AD clinical staff which includes a psychiatrist, neurologist, and neuropsychologist. Familial EOAD cases were defined as AAO <65 years of age. As reported in previous studies, AAO was defined as the age at which an individual or family historian reported onset of significant cognitive problems that interfered with normal activity, or the AAO of problems as documented in the medical record.[15] All affected individuals met the internationally recognized standard NINCDS-ADRDA criteria.[16,17] The cognitive status of participants was measured using either the Mini-Mental State Examination[18] or the Modified Mini-Mental State.[19]

Patients with LOAD (N = 151) were part of a study investigating coding mutations in *SORL1* in AD.[11] These participants were drawn from a larger study of AD genetics restricted to Caribbean Hispanics. All affected individuals were of Caribbean Hispanic ancestry. All participants were assessed using standard clinical examinations and cognitive testing as described elsewhere.[20] For this study, we reviewed the clinical records of participants who had *SORL1* mutations to assess for possible features of Parkinson disease (PD) or more broadly, parkinsonism.

WES and variant calling. All samples were prepared using DNA extracted from the blood. Genomic DNA was then sheared and processed using the SureSelect Human All Exon 50 Mb v4 capture kit (Agilent Technologies, Santa Clara, CA) according to the manufacturer's protocol at the HIHG Center for Genome Technology. After capture, the DNA was tested for uniform enrichment of targets via quantitative PCR. Sequencing was then performed on the Illumina HiSeq2000 at 2× 150 bp paired-end cycles at 40–50× on target depth. Exomes were sequenced to sufficient depth to achieve a minimum threshold of 80% of coding sequence covered with at least 15 reads, based on UCSC hg19 "known gene" transcripts. The mean depth of coverage across *SORL1* was 68.

Sequencing data from the Illumina HiSeq2000 were processed using an established semiautomated pipeline. Initial image files were processed using the Firecrest module (Illumina, San Diego, CA) to determine cluster intensities and noise. After initial quality control, BWA-ELAND and CASAVA v1.9 were used for realignment to the human genome version hg19. Results from BWA and CASAVA are then fed into additional software packages (CLC Genomics Workbench and GenomeStudio) for secondary analysis, visualization of the called variants, and browsing of consensus reads.[21,22] Genotype calling was performed with GATK Unified Genotyper. Variants were then normalized using BCFTools.[23] Single nucleotide polymorphisms with read depth <6, variant quality score log odds ratio <0, and Phred-scaled likelihood score <100 were removed from further analysis. Variants were filtered to identify alterations that were likely to be damaging (missense, splicing, stop-gain, stop-loss, and insertion/deletions) in Gencode v19, NCBI RefSeq, or Ensembl gene annotations.[24,25] Variants were screened to determine whether they occurred in a known or suspected EOAD gene (*APP*, *GRN*, *MAPT*, *PSEN1*, *PSEN2*, *SORL1*, and *TREM2*). Minor allele frequencies were obtained from the Exome Aggregation Consortium.[26]

Cloning of *SORL1* variants. Site-directed mutagenesis was used to generate the SORL1 T588I and SORL1 T2134IM mutation constructs using human SORL1-MYC pcDNA3.1 as a backbone according to the manufacturer's instructions as previously published.[3,11,27–30] Sequencing was used to verify mutant constructs. Cell culture and transfection followed previously described standard protocols.[3,11,27–30]

Aβ, Western blot, and co-immunoprecipitation assays. Aβ assays were measured by sandwich ELISA assay in culture medium from stably transfected HEK293 cells expressing the Swedish APP mutant (APPsw) and either wild-type SORL1 or mutant SORL1 as previously described.[3,11,27–30] Cell surface biotinylation was performed using 1 mg/mL Sulfo-NHS-LC-Biotin (Sigma-Aldrich, St. Louis, MO) for 20 minutes at 4°C to prevent internalization. Cells were then washed and lysed, and biotinylated proteins were precipitated with NeutrAvidin beads (Thermo Fisher Scientific, Waltham, MA). Western blot band intensities were measured with ImageJ software and samples normalized to the wild-type control. Co-immunoprecipitation was performed after cell lysis in 1% CHAPSO buffer,[3] using G Plus beads with 2 μg mouse monoclonal anti-c-MYC antibody for the immunoprecipitation of SORL1-myc, immunoblotted with anti-C-terminal APP antibody (Ab365), and anti-C-terminal SORL1 (S9200). Western blot band intensities were

measured with ImageJ software. Full-length (FL) APP coprecipitated with c-MYC antibody was quantified and normalized to the amount of immunoprecipated SORL1 as previously described.[3,11,27–30]

Statistical analyses. Statistical analyses were performed using Graphpad statistical software (graphpad.com/guides/prism/5/user-guide/prism5help.html?using_tour_overview.htm; GraphPad Prism 5). Analysis of variance and *t* tests were used to analyze statistical difference, followed by Bonferroni correction (*$p < 0.05$; **$p < 0.01$; and ***$p < 0.001$).

RESULTS *SORL1* variants in EOAD families. WES identified 10 individuals with *SORL1* mutations in 2 unrelated EOAD families (table 1, figure 1). Neuropathology results were available for 1 affected individual. The first family, number 191, has 6 individuals with the predicted damaging SORL1 T588I mutation (rs752726649; C>T); all 4 affected individuals for whom DNA was available were found to carry this variant. These 4 affected individuals had AAOs that ranged from 59 to 82 years. While the progressive cognitive decline of each individual was consistent with dementia, individuals 104 and 111 had also parkinsonian features. Individual 104 began to show cognitive impairment at age 82. On examination, he demonstrated tremor at rest, hypophonia, micrographia, masked facial expression, smaller steps on gait, and overall bradykinesia. Chart review indicated that these symptoms were levodopa/carbidopa responsive. Imaging revealed white matter changes and moderate cerebral atrophy, and EEG was remarkable for a loss of alpha waves. Individual 111 had the earliest AAO in the family 191 at age 59, with diminished memory function in all domains, clinically judged to most likely represent EOAD. When seen by research staff at age 70, the individual was noted to exhibit parkinsonian features. This presentation was confounded by several years of treatment with haloperidol, a typical antipsychotic agent that can cause parkinsonian side effects. Two unaffected individuals in family 191 also carried the SORL1 T588I mutation. These individuals were last examined at ages 81 and 84 years, respectively. Individuals 116 and 9004 demonstrated a normal cognitive and physical examination.

The second family, number 1240 (table 1 and figure 1), contains 3 affected individuals with the SORL1 T2134M mutation (rs142884576; C>T). These 3 affected individuals had AAOs that ranged from 55 to 84 years. While the clinical examinations revealed no motor abnormalities, there was autopsy evidence for Lewy bodies in individual 1, with the earliest AAO in the family at 55 years. Neuropathologic diagnosis of individual 1 was indicative of Braak & Braak stage IV tangles and limbic Lewy bodies. In addition, 1 individual (119) demonstrated progressive cognitive decline consistent with AD without the T2134M SORL1 mutation. This individual had an AAO of 76 years. Finally, there was 1 unaffected individual (113) with this T2134M SORL1 mutation who was last examined at 79 years of age.

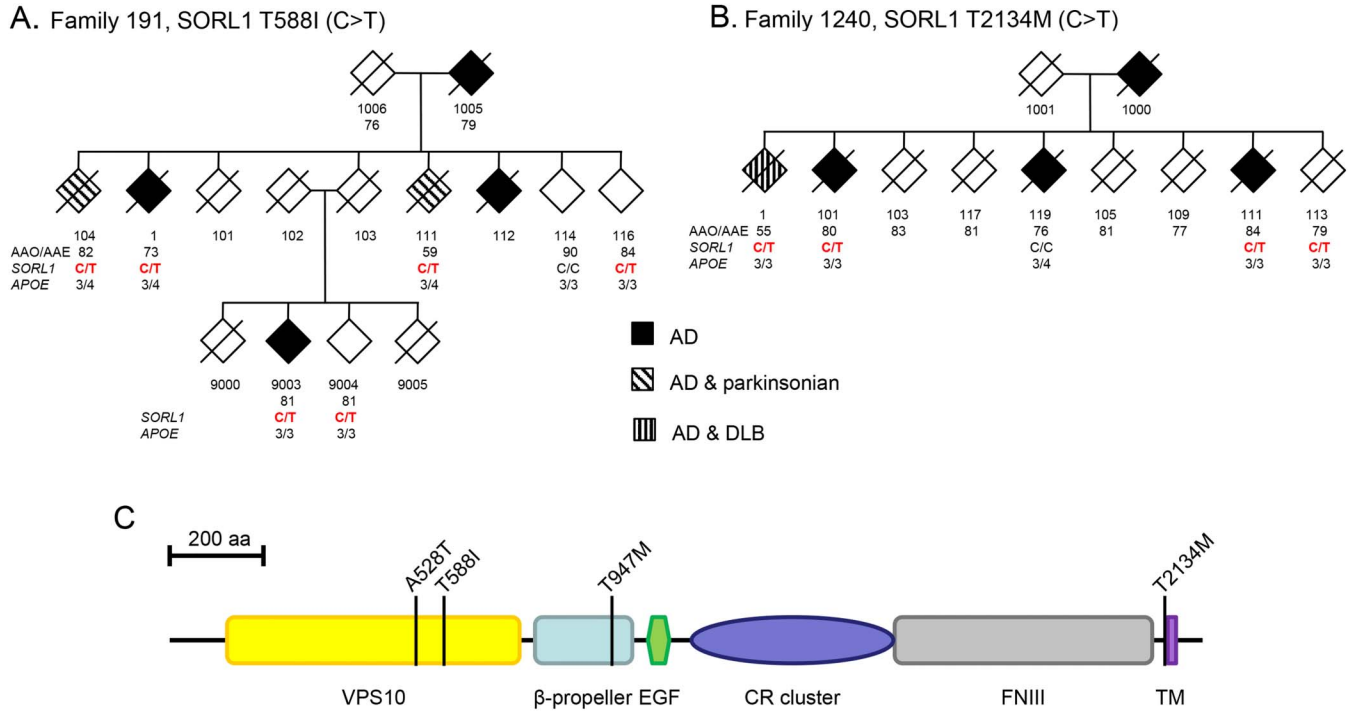
Parkinsonian features in patients with LOAD with *SORL1* variants. Given the clinical results from these 2 EOAD families, we examined in greater depth the clinical status of previously reported patients with *SORL1* changes.[11] Review of clinical history and physical examination data identified 4 additional AD individuals, all with LOAD (no neuropathology results were available), and who had evidence of

Table 1 Clinical features in early-onset AD families with *SORL1* variants

Family-individual	Sex	Affection status	Additional features	AAO/AAE	Base pair position (hg38 assembly)	Nucleotide	Amino acid	dbSNP number[a]	*APOE* status
191-1	M	Dementia AD	—	73	121543625	C>T	T588I	rs752726649	3/4
191-104	F	Dementia AD	Parkinsonian	82	121543625	C>T	T588I	rs752726649	3/4
191-111	M	Dementia AD	Parkinsonian	59	121543625	C>T	T588I	rs752726649	3/4
191-114	M	—	—	90	—	—	—	—	3/3
191-116	M	—	—	84	121543625	C>T	T588I	rs752726649	3/3
191-9003	M	Dementia AD	—	81	121543625	C>T	T588I	rs752726649	3/3
191-9004	M	—	—	81	121543625	C>T	T588I	rs752726649	3/3
1240-1	M	Dementia AD	DLB, no PD	55	121627591	C>T	T2134M	rs142884576	3/3
1240-101	F	Dementia AD	—	80	121627591	C>T	T2134M	rs142884576	3/3
1240-111	F	Dementia AD	—	84	121627591	C>T	T2134M	rs142884576	3/3
1240-113	M	—	—	79	121627591	C>T	T2134M	rs142884576	3/3
1240-119	F	Dementia AD	—	76	—	—	—	—	3/4

Abbreviations: AAE = age at examination; AAO = age at onset; AD = Alzheimer disease; DLB = dementia with Lewy bodies; PD = Parkinson disease; SNP = single nucleotide polymorphism.
[a] Minor allele frequencies (MAF): rs752726649 global MAF = 8.2×10^{-6}; rs142884576 global MAF = 2.2×10^{-4}.

Figure 1 **Pedigrees of the early-onset Alzheimer disease families and SORL1 protein diagram**



(A and B) Pedigrees of the early-onset Alzheimer disease (EOAD) families and SORL1 protein diagram. Affected individuals are solid black while those presenting with parkinsonian features are patterned. Below each individual number is either the age at onset (AAO, for affected individuals) or the age at last examination (AAE, for unaffected individuals). For family 191, the *SORL1* variant is present in all affected individuals examined. In family 1240, the variant occurs in 3 of 4 cases evaluated. (C) Diagram of SORL1 protein (2214 total amino acids) indicating the location of principal domains and the variants identified in the EOAD families (T588I and T2134M) and the late-onset Alzheimer disease individuals (A528T and T947M). AD = Alzheimer disease; VPS10 = vesicular protein sorting 10 domain; CR = complement type repeat domains; EGF = epidermal growth factor; FNIII = fibronectin type III repeats; TM = transmembrane region; DLB = dementia with Lewy bodies; *SORL1* = sortilin-related receptor.

parkinsonian features (table 2). The *SORL1* mutations in these 4 individuals were distinct from those identified in the first 2 families. Specifically, 3 individuals which we previously reported carry a common variant at A528T (rs2298813A>G). Clinically, these individuals were diagnosed with both AD and PD and had ages of AD onset ranging from 78 to 84 years. The fourth individual had a different previously reported missense T947M variant (rs143571823, C>T). This individual had a clinical diagnosis of AD and parkinsonism with an age of AD onset at 90 years.

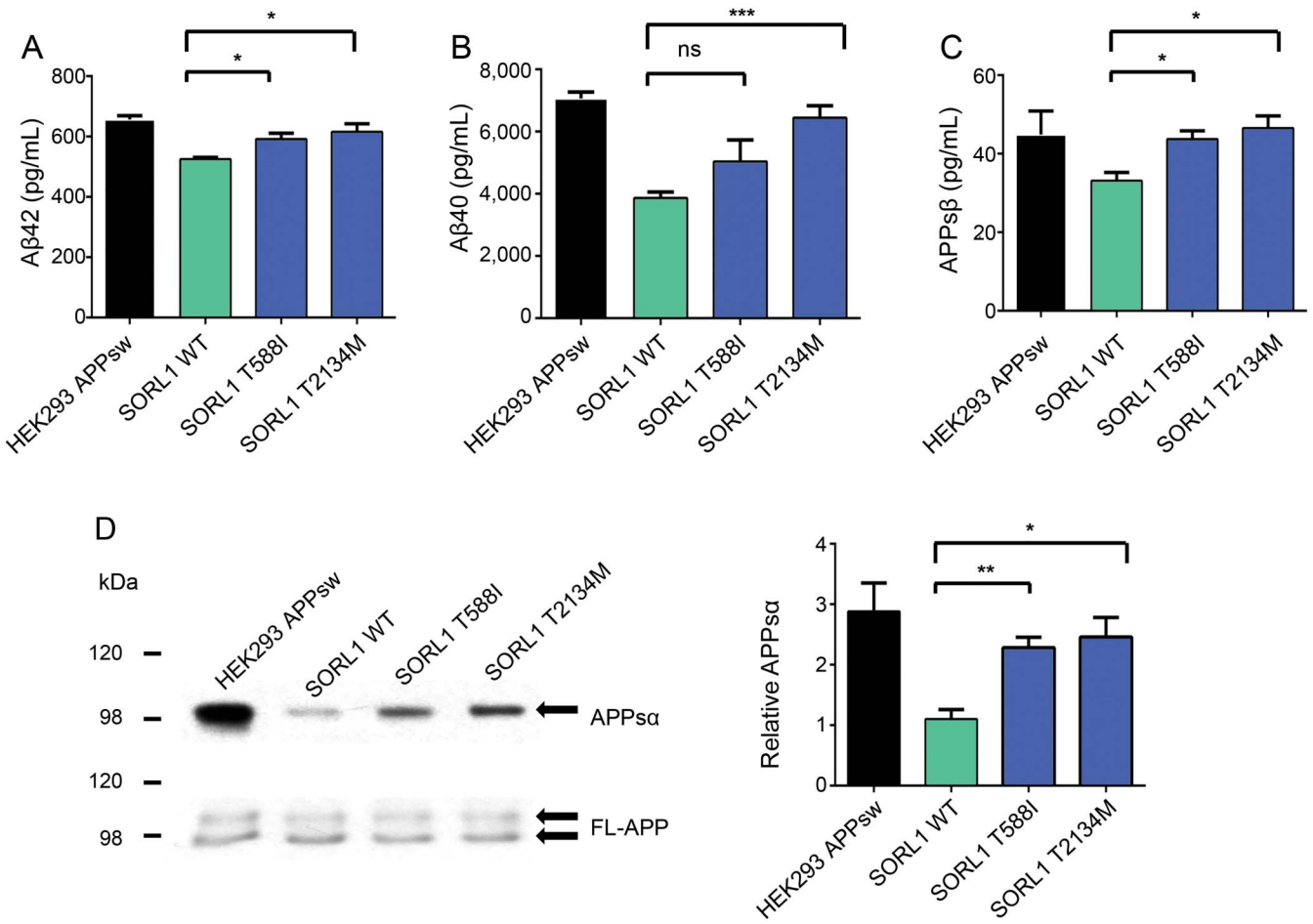
***SORL1* variants alter Aβ levels and APP trafficking.** Next, we examined the functional consequences of the SORL1 T588I and T2134M alterations identified in the EOAD families; the variants identified in the LOAD individuals (A528T and T947M) were previously assessed and reported.[11] To determine the effects on Aβ production by these *SORL1* variants, Aβ42 and Aβ40 levels were measured in conditioned media collected from cultured HEK293 cells expressing equivalent levels of wild-type SORL1 protein, SORL1 T588I, or SORL1 T2134M. Both mutants increased Aβ42 secretion compared with the control (T588I: 113% ± 1.6% and T2134M: 117% ± 5.1%, $p < 0.05$, figure 2A). Overexpression of SORL1 T2134M also increased Aβ40 secretion (167% ± 9.9%, $p < 0.001$, figure 2B). While the

Table 2 **Clinical features in late-onset AD individuals with *SORL1* variants**

Family-individual	Sex	Affection status	Additional features	AAO	Base pair position (hg38 assembly)	Nucleotide	Amino acid	dbSNP number[a]	*APOE* status
216-9	M	Dementia AD	Parkinsonian	78	121522975	A>G	A528T	rs2298813	2/4
920-10	F	Dementia AD	Parkinsonian	83	121522975	A>G	A528T	rs2298813	3/4
1280-18	F	Dementia AD	Parkinsonian	90	121558767	C>T	T947M	rs143571823	3/4
1731-1	M	Dementia AD	Parkinsonian	84	121522975	A>G	A528T	rs2298813	3/3

Abbreviations: AAO = age at onset; AD = Alzheimer disease; SNP = single nucleotide polymorphism.
[a] Minor allele frequencies (MAF): rs2298813 MAF = 0.072; rs143571823 MAF = 0.0013.

Figure 2 ***SORL1* mutants' overexpression increases β-amyloid secretion**



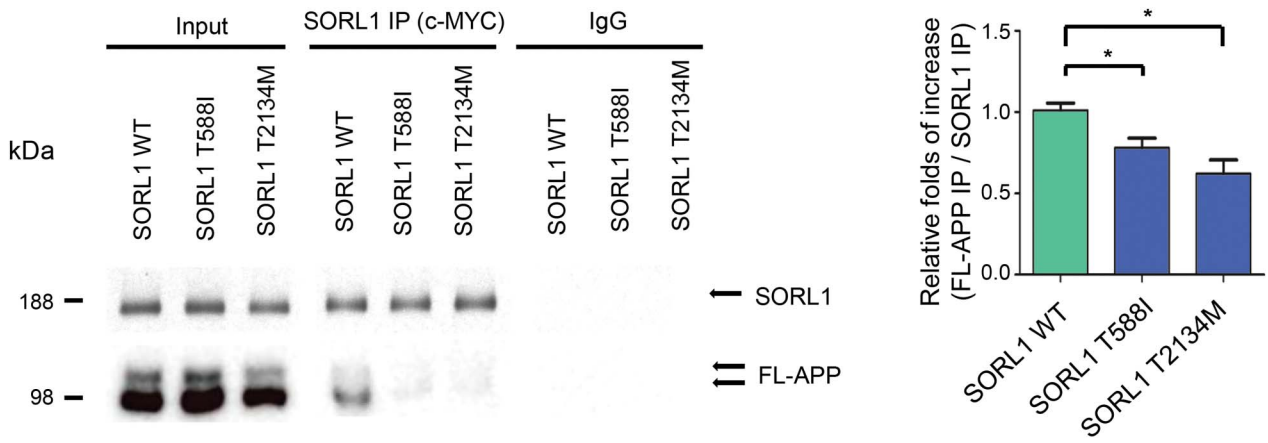
(A–C) Secreted β-amyloid 40 (Aβ40), Aβ42, and amyloid precursor protein soluble β-secretase (APPsβ) were measured from culture medium in stable HEK293 cells expressing the APP Swedish mutant (HEKsw) together with either wild-type *SORL1* or mutant *SORL1*. Error bars represent standard error of the mean (SEM). ***$p < 0.001$, *$p < 0.05$, ns, not significant, n = 3 independent replications. (D) Western blot was performed to detect APP soluble α-secretase (APPsα) from cultured media. Bar graphs were normalized to control. **$p < 0.01$, n = 3 independent replications, and error bars represent the SEM. Aβ = β-amyloid; FL-APP = full-length APP; *SORL1* = sortilin-related receptor.

SORL1 T588I alteration trended toward an increase of Aβ40 secretion in cells, it did not reach statistical significance (131% ± 17.6%, figure 2B).

SORL1 has been proposed to modulate the post-translational biology of APP at several intracellular sites including during transport out of the Golgi and during re-entry and recycling from the cell surface. To examine further the effect of these *SORL1* mutants on APP trafficking, we measured APP soluble β-secretase (APPsβ) secretion in a conditioned medium.[31] Both mutations caused an increase in APPsβ secretion compared with the wild-type *SORL1* (T588I: 132% ± 6.3%, $p < 0.05$; T2134M: 140% ± 9.4%, $p < 0.05$, figure 2C). Both mutations also increased production of the soluble α-secretase cleavage product compared with control cells (T588I: 207% ± 15.8%, $p < 0.01$; T2134M: 223% ± 29.6%, $p < 0.05$, figure 2D). These observations suggest that in the presence of these *SORL1* mutants, APP is neither retained efficiently in the Golgi nor effectively retrieved from the cell surface into recycling pathways. This could result in additional APP lingering at the cell surface. This hypothesis was supported by surface biotinylation experiments which revealed that both *SORL1* mutants increased the amount of surface APP compared with the control (T588I: 143% ± 13.1%, $p < 0.05$; T2134M: 138% ± 7.5%, $p < 0.05$, figure e-1 at Neurology.org/ng).

***SORL1* variants decrease APP binding.** To understand the mechanism by which these *SORL1* mutants might alter APP trafficking at the cell surface, we next measured levels of SORL1 protein at the cell surface. The T588I variant showed essentially normal levels of SORL1 both at the cell surface and in total cell lysates (~87% ± 13.1% of control value, figure e-1). However, while the T2134M mutant showed normal levels of total cellular SORL, there were decreased amounts of surface SORL1 (~25%, $p < 0.05$, figure e-1B).

Figure 3 **Both *SORL1* mutants have a reduced binding affinity to APP**



SORL1 was immunoprecipitated from cell lysates with a c-MYC antibody, and the amount of coprecipitated full-length amyloid precursor protein (FL-APP) was measured by densitometry of the anti-APP immunoreactive bands on the Western blot of the SORL1 immunoprecipitation products. *$p < 0.05$, **$p < 0.01$, n = 3 replications, and error bars represent the SEM. IgG = immunoglobulin G; IP = immunoprecipitated; *SORL1* = sortilin-related receptor.

Previous work by us and others have demonstrated that SORL1 directly binds APP and regulates its sorting into secretory, endocytic, or recycling pathways.[3,4,11,27,32–38] To assess whether the SORL1 T588I and T2134M mutations might alter the binding affinity of SORL1 to APP, we immunoprecipitated SORL1 from whole cell lysates using an anti-myc antibody directed to the myc epitope on the exogenous SORL1 protein. This strategy circumvents possible risk that the *SORL1* mutants might alter binding affinity of anti-SORL1 antibodies, or that endogenous SORL1 might be pulled down in addition, to overexpressed SORL1 in the mutant APPsw cell lines. We then measured the amount of FL-APP that co-immunoprecipitated with the myc-tagged SORL1 proteins and expressed the binding as a normalized ratio of the abundance of coprecipitated FL-APP relative to the abundance of immunoprecipitated SORL1. Both mutations caused reductions in APP binding (T588I: ~77.1% ± 5.8%, $p < 0.05$; T2134M: ~61.5% ± 8.3%, $p < 0.05$, figure 3).

DISCUSSION In this study, we identified *SORL1* alterations in EOAD families thus confirming previously reported studies showing a role for *SORL1* in risk for EOAD. Furthermore, we presented functional evidence that these *SORL1* alterations are pathogenic.

Evidence for functional consequences of *SORL1* mutations is scant. However, the evidence shown here suggests that the variants identified in the EOAD families, SORL1 T588I and T2134, weaken the interaction of SORL1 with FL-APP. This can culminate in excessive APP accumulating at the cell surface either due to failure of the mutant SORL1 to slow trafficking of APP to the cell surface[39] or failure of mutant SORL1 to retrieve FL-APP into the retromer-recycling endosome pathway.[3,4,11,27,32–38] Our result agrees with prior work which suggests that some *SORL1* mutants cause reduced trafficking of the mutant SORL1 protein from the endoplasmic reticulum (ER)/Golgi network to the cell surface.[11] The resulting misdirection of more APP into the late endosome pathway exposes the APP to β-secretase and γ-secretase cleavage, with the consequent increase in Aβ production, especially Aβ42. Intriguingly, but consistent with prior work, our data suggest that the molecular mechanisms underlying this common overall effect differ between the 2 variants. Thus, the T2134M mutant, which is located close to the transmembrane domain (figure 1), appears to disrupt trafficking of SORL1 to the cell surface, presumably due to its removal from the ER-Golgi secretory pathway by the ER quality control systems which remove misfolded proteins. In contrast, the T588I mutant survives the ER quality control mechanisms, but appears to be less efficient than wild-type SORL1 in binding to APP. The molecular mechanism for the reduced binding of T588I is unclear, but may relate to subtle changes in the fold of the extracellular domain of SORL1 such that putative APP-binding sites in VPS10 and/or in complement type repeat domains.[39,40] Crucially, while they may have different underlying molecular mechanisms, the net effect of both mutations is the same.

A secondary finding in our study was the observation of additional clinical features beyond AD among select individuals with *SORL1* alterations. These clinical findings, based on extensive clinical reviews, included clinical Parkinson-related features and neuropathology-proven Lewy bodies without clinical parkinsonism. While these findings point to

a potential association between *SORL1* alterations and a broader spectrum of neurodegenerative disorders, it is important to note that these clinical features were not present in all individuals with *SORL1* alterations and may simply represent features of coincidental sporadic PD.

The results from this study demonstrate that select *SORL1* variants present in EOAD and LOAD alter Aβ levels and interfere with APP trafficking. In addition, we observed parkinsonian features among some EOAD/LOAD individuals with *SORL1* alterations. These clinical findings should be viewed cautiously but suggest the need for exploration of the additional phenotype consequences of *SORL1* alterations beyond dementia.

AUTHOR CONTRIBUTIONS

M.L.C., R.M.C., B.W.K., and M.A.P.-V. conceived and designed the experiments. M.L.C., R.M.C., R.M., and M.A.P.-V. acquired and assessed participants. P.L.W. and H.N.C. performed custom capture sequencing and exome sequencing. B.W.K., H.N.C., and M.A.P.-V. analyzed the sequencing data. B.W.K. and M.A.P.-V. performed the statistical analysis. Y.Z., C.B., and P.S.G.-H. cloned the *SORL1* variants, performed all assays, and analyzed all resulting data. M.L.C. and R.M.C. drafted the manuscript. M.L.C., R.M.C., Y.Z., C.B., B.W.K., H.N.C., P.S.G.-H., and M.A.P.-V. edited the manuscript. The authors jointly discussed the experimental results over the course of the study. All authors read and approved the final manuscript.

ACKNOWLEDGMENT

The authors are grateful to the families and staff who participated in this study. In addition, they acknowledge Larry Deon Adams for his assistance in data management and Anthony Griswold for his help in bioinformatics.

STUDY FUNDING

This research was supported by grants from the National Institutes of Health (R01 AG027944 and R01 AG028786 to M.A.P.-V.; R01 AG019085 to J.L.H.; P20 MD000546 and R01 AG28786-01A1 to G.S.B.; U01-AG032984 and RC2-AG036528 to G.D.S., U24 AG026395, U24 AG026390, R01AG037212, R37 AG015473, and R01AG015473 to R.M.; R01 AG009029 to L.A.F.; U01-AG016976 to W.A.K.; U24-AG021886 to T.M.F.; R01AG009956 and RC2 AG036650 to K.H.; UO1 AG06781 and UO1 HG004610 to E.L.; 5R01AG20688 to M.D.F.; P50 AG005133 and AG030653 to I.K.; R01 AG1101, R01 AG030146, and RC2 AG036650 to D.A.E.; P30AG10161, R01AG15819, R01AG30146, R01AG17917, and R01AG15819 to D.A.B.; R01AG028786 to J.J.M.; R01AG22018 and P30AG10161 to L.L.B.; P50AG16574, R01 032990, and KL2 RR024151 to N.E.T. and N.R.G.R.; AG005138 to J.D.B.; P50 AG05681, P01 AG03991, and P01 AG026276 to A.M.G.; U24-AG041689 to L.S.W.; and U19 AG047133 and UF1AG047133); the Department of Defense (W81XWH-12-1-0013 to M.A.P.-V.); a joint grant from the Alzheimer's Association (SG-14-312644) and the Fidelity Biosciences Research Initiative to M.A.P.-V.; and the BrightFocus Foundation (A2011048 to M.A.P.-V.).

DISCLOSURE

Dr. Cuccaro has served on the editorial board of *Child Psychiatry & Human Development*. Dr. Carney, Dr. Zhang, Dr. Bohm, and Dr. Kunkle report no disclosures. Dr. Vardarajan has served on the scientific advisory board of, been a consultant of, and received research support from the Immuneering Corporation. Ms. Whitehead reports no disclosures. Dr. Cukier has served on the scientific advisory board of the General Direction for Scientific Research and Health Innovation (Italian Ministry of Health); and has received research support from NARSAD, the Hussman Foundation, and the Alzheimer's Association. Dr. Mayeux has received research support from the following grants: RO1 AG037212 Epidemiology of Biomarkers of Risk and Progression in Late onset Alzheimer's Disease; and RF1 AG015473 Genetic Studies of Alzheimer's Disease in Caribbean Hispanics. Dr. St. George-Hyslop has served on the scientific advisory board of the University of Lille (basic AD program) and has received research support from the Canadian Institutes of Health, the Medical Research Council, and Wellcome Trust. Dr. Pericak-Vance reports no disclosures. Go to Neurology.org/ng for full disclosures.

Received June 13, 2016. Accepted in final form September 23, 2016.

REFERENCES

1. Karch CM, Goate AM. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. Biol Psychiatry 2015;77:43–51.
2. Alzheimer's Association. 2015 Alzheimer's disease facts and figures. Alzheimers Dement 2015;11:332–384.
3. Rogaeva E, Meng Y, Lee JH, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet 2007;39:168–177.
4. Offe K, Dodson SE, Shoemaker JT, et al. The lipoprotein receptor LR11 regulates amyloid beta production and amyloid precursor protein traffic in endosomal compartments. J Neurosci 2006;26:1596–1603.
5. Lee JH, Barral S, Reitz C. The neuronal sortilin-related receptor gene SORL1 and late-onset Alzheimer's disease. Curr Neurol Neurosci Rep 2008;8:384–391.
6. Meng Y, Lee JH, Cheng R, St George-Hyslop P, Mayeux R, Farrer LA. Association between SORL1 and Alzheimer's disease in a genome-wide study. Neuroreport 2007;18:1761–1764.
7. Kolsch H, Jessen F, Wiltfang J, et al. Association of SORL1 gene variants with Alzheimer's disease. Brain Res 2009;1264:1–6.
8. Lee JH, Cheng R, Honig LS, Vonsattel JP, Clark L, Mayeux R. Association between genetic variants in SORL1 and autopsy-confirmed Alzheimer disease. Neurology 2008;70:887–889.
9. Yin RH, Yu JT, Tan L. The role of SORL1 in Alzheimer's disease. Mol Neurobiol 2015;51:909–918.
10. Reitz C, Cheng R, Rogaeva E, et al. Meta-analysis of the association between variants in SORL1 and Alzheimer disease. Arch Neurol 2011;68:99–106.
11. Vardarajan BN, Zhang Y, Lee JH, et al. Coding mutations in SORL1 and Alzheimer disease. Ann Neurol 2015;77:215–227.
12. Pottier C, Hannequin D, Coutant S, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. Mol Psychiatry 2012;17:875–879.
13. Nicolas G, Charbonnier C, Wallon D, et al. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. Mol Psychiatry 2016;21:831–836.
14. 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.
15. Carney RM, Slifer MA, Lin PI, et al. Longitudinal follow-up of late-onset Alzheimer disease families. Am J Med Genet B Neuropsychiatr Genet 2008;147B:571–578.
16. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. Neurology 1984;34:939–944.

17. 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.
18. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–198.
19. Teng EL, Chui HC. The modified mini-mental state (3MS) examination. J Clin Psychiatry 1987;48:314–318.
20. 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.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25:1754–1760.
22. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20: 1297–1303.
23. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27:2987–2993.
24. Raney BJ, Dreszer TR, Barber GP, et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC genome browser. Bioinformatics 2014; 30:1003–1005.
25. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res 2002;12:996–1006.
26. Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–291.
27. Reitz C, Tokuhiro S, Clark LN, et al. SORCS1 alters amyloid precursor protein processing and variants may increase Alzheimer's disease risk. Ann Neurol 2011;69: 47–64.
28. Bohm C, Seibel NM, Henkel B, Steiner H, Haass C, Hampe W. SorLA signaling by regulated intramembrane proteolysis. J Biol Chem 2006;281:14547–14553.
29. Hasegawa H, Sanjo N, Chen F, et al. Both the sequence and length of the C terminus of PEN-2 are critical for intermolecular interactions and function of presenilin complexes. J Biol Chem 2004;279:46455–46463.
30. Yu G, Nishimura M, Arawaka S, et al. Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing. Nature 2000;407:48–54.
31. Vassar R, Bennett BD, Babu-Khan S, et al. Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. Science 1999; 286:735–741.
32. Jacobsen L, Madsen P, Jacobsen C, Nielsen MS, Gliemann J, Petersen CM. Activation and functional characterization of the mosaic receptor SorLA/LR11. J Biol Chem 2001;276: 22788–22796.
33. Andersen OM, Reiche J, Schmidt V, et al. Neuronal sorting protein-related receptor sorLA/LR11 regulates processing of the amyloid precursor protein. Proc Natl Acad Sci USA 2005;102:13461–13466.
34. Dodson SE, Andersen OM, Karmali V, et al. Loss of LR11/SORLA enhances early pathology in a mouse model of amyloidosis: evidence for a proximal role in Alzheimer's disease. J Neurosci 2008;28:12877–12886.
35. Hermey G. Targeting amyloid precursor protein. Ann Neurol 2011;69:8–10.
36. Fjorback AW, Seaman M, Gustafsen C, et al. Retromer binds the FANSHY sorting motif in SorLA to regulate amyloid precursor protein sorting and processing. J Neurosci 2012;32:1467–1480.
37. Lane RF, Steele JW, Cai D, Ehrlich ME, Attie AD, Gandy S. Protein sorting motifs in the cytoplasmic tail of SorCS1 control generation of Alzheimer's amyloid-beta peptide. J Neurosci 2013;33:7099–7107.
38. Schmidt V, Sporbert A, Rohe M, et al. SorLA/LR11 regulates processing of amyloid precursor protein via interaction with adaptors GGA and PACS-1. J Biol Chem 2007; 282:32956–32964.
39. Mehmedbasic A, Christensen SK, Nilsson J, et al. SorLA complement-type repeat domains protect the amyloid precursor protein against processing. J Biol Chem 2015;290: 3359–3376.
40. Gustafsen C, Glerup S, Pallesen LT, et al. Sortilin and SorLA display distinct roles in processing and trafficking of amyloid precursor protein. J Neurosci 2013;33:64–71.

Neuroscience Letters 649 (2017) 124–129

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Research article

Targeted sequencing of *ABCA7* identifies splicing, stop-gain and intronic risk variants for Alzheimer disease

B.W. Kunkle[a], R.M. Carney[a,b], M.A. Kohli[a], A.C. Naj[a,c], K.L. Hamilton-Nelson[a], P.L. Whitehead[a], L. Wang[a], R. Lang[d], M.L. Cuccaro[a], J.M. Vance[a], G.S. Byrd[d], G.W. Beecham[a], J.R. Gilbert[a], E.R. Martin[a], J.L. Haines[e], M.A. Pericak-Vance[a,*]

[a] John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA
[b] Miami Veterans' Affairs Medical Center, Miami, FL, USA
[c] Department of Biostatistics & Epidemiology, University of Pennsylvania, Philadelphia, PA, USA
[d] Department of Biology, North Carolina A & T University, Greensboro, NC, USA
[e] Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA

HIGHLIGHTS

- Sequencing of the Alzheimer disease risk locus *ABCA7* is performed.
- Several Alzheimer's disease risk variants are identified in the gene *ABCA7*.
- Three previously associated *ABCA7* variants are confirmed.
- A 3′-UTR splice variant in ABCA7 is identified as a potential risk variant.

ARTICLE INFO

Article history:
Received 29 December 2016
Accepted 7 April 2017
Available online 8 April 2017

Keywords:
Alzheimer
Genetics
Sequencing
ABCA7
Splicing
Intronic

ABSTRACT

Several variants in the gene *ABCA7* have been identified as potential causal variants for late-onset Alzheimer's disease (LOAD). In order to replicate these findings, and search for novel causal variants, we performed targeted sequencing of this gene in cohorts of non-Hispanic White (NHW) and African-American (AA) LOAD cases and controls. We sequenced the gene *ABCA7* in 291 NHW LOAD cases and 103 controls. Variants were prioritized for rare, damaging variants and previously reported variants associated with LOAD, and were follow-up genotyped in 4076 NHW and 1157 AA cases and controls. We confirm three previously associated *ABCA7* risk variants and extend two of these associations to other populations, an intronic variant in NHW ($P = 3.0 \times 10^{-3}$) (originally reported in a Belgian population), and a splice variant originally associated in the Icelandic population, which was significantly associated in the NHW cohort ($P = 1.2 \times 10^{-6}$) and nominally associated in the AA cohort ($P = 0.017$). We also identify a 3′-UTR splice variant that segregates in four siblings of one family and is nominally associated with LOAD ($P = 0.040$). Multiple variants in *ABCA7* contribute to LOAD risk.

1. Introduction

Recent genome-wide association studies (GWAS) have identified common variants in *ABCA7* as associated with late-onset Alzheimer's disease (LOAD) [1–4]. The common single nucleotide polymorphisms (SNPs) associated in these studies confer only modest risk, and have no known or apparent functional consequences related to development of LOAD. However, recent sequencing studies of LOAD loci have identified several potential causal variants in *ABCA7*, including intronic, missense, and frameshift variants [5–8].

To identify additional low-frequency and rare variants that are potentially causal in *ABCA7*, and to confirm and examine the generalizability of previously reported candidate risk variants within *ABCA7*, we performed targeted sequencing of this gene in a discovery set of 291 LOAD cases and 103 cognitively intact controls. Candidate causal variants from this sequencing, along with previously associated rare (minor-allele frequency (MAF) $\leq$ 0.01) and

* Corresponding author at: John P Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, 1501 NW 10th Ave, Miami, FL 33136, USA.
E-mail addresses: m.pericakvance@miami.edu, mpericak@med.miami.edu (M.A. Pericak-Vance).

http://dx.doi.org/10.1016/j.neulet.2017.04.014
0304-3940/

Table 1
Characteristics of cases and controls selected for targeted sequencing.

	Cases	Controls
Total	291	103
Female, n (%)	178 (61.4%)	46 (44.7%)
Age (years), mean [SD]	72.4 [7.8]	83.6 [3.4]
APOE genotype		
ε3/ε3, n	247	103
ε3/ε4, n	43	–

low-frequency (MAF > 0.01 and ≤ 0.05) variants in this locus, were identified for follow-up association testing in independent familial and case-control datasets. Identification of functional variants from these analyses and similar studies could prove important for development of therapeutics targeting *ABCA7*.

2. Materials and methods

2.1. Targeted sequencing and follow-up genotyping sample selection

All cases and controls selected for targeted sequencing were from The John P. Hussman Institute for Human Genomics (HIHG) at the University of Miami and Case Western Reserve University (CWRU) Alzheimer disease cohort. The HIHG/CWRU cohort contains 1270 NHW cases and 1661 cognitively healthy NHW controls, of which 1574 individuals are from 511 LOAD families. Patients were collected over the course of ~30 years, with protocols and amendments being approved at each stage. Across the multiple sites, cognitive status of controls was measured using the MMSE/3MS and Clinical Dementia Rating Scale. Diagnostic criteria followed that of the National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) for probable or definite AD with age at onset greater than 60 [9,10]. An extreme-sampling strategy was used for selection of 291 cases and 103 controls for targeted sequencing from this cohort. Cases were *APOE* ε3/ε3 or *APOE* ε3/ε4 and age at onset >65, while all controls were *APOE* ε3/ε3 with an age at last exam >80. Cognitively healthy controls were unrelated individuals from the same geography, were frequency matched by age and sex, and had a documented MMSE or 3MS score in the normal range. Mean ages-at-onset (AAO)/ages-at-exam (AAE) for cases and controls were 74 (±8 standard deviations (SD)) and 83.6 (±3.4 SD), respectively. Cases are 63% female, and controls are 61% female (Table 1).

Follow-up genotyping was conducted in 4076 NHW cases and controls and 1157 African American (AA) cases and controls. The NHW participants are part of three different cohorts: 1) The HIHG/CWRU Cohort, 2) The National Institute of Mental Health (NIMH) Genetic Studies of Alzheimer's Disease Cohort, and 3) The National Cell Repository Alzheimer Disease (NCRAD)/National Institute on Aging Late-Onset Alzheimer Disease (NIA-LOAD) Family Study. All participants in these cohorts were enrolled following informed consent and using protocols approved by the appropriate Institutional Review Boards. A breakdown of cases and controls per cohort can be found in Supplementary Table 1. All individuals enrolled self-identified as NHW and ethnicity was confirmed using high density SNP data analysis [2,11,12].

The NIMH Cohort [13,14] is a publicly-available sample containing LOAD pedigrees ascertained by three sites (University of Alabama – Birmingham, Johns Hopkins University, and Massachusetts General Hospital/Harvard Medical School). Our analysis included 822 cases and 357 unaffected individuals from 397 pedigrees. Each pedigree has at least two affected individuals who are biologically related as first-, second-, or third- degree (first cousins only) relatives. The NCRAD/NIA-LOAD study included 186 affected and 174 unaffected individuals from 232 families. The study recruited families with two or more affected siblings with LOAD and unrelated controls matched for age and ethnic background. Further details of the study recruitment and cognitive assessment procedures have been previously described [15].

All AA cases and controls are from the African American Alzheimer's Disease Genomics Coalition (AAADGC) from three contributing sites including the HIHG, CWRU and North Carolina A&T University, Greensboro, NC. This dataset contains 484 AA cases (370 women and 114 men, mean AAO 74.0 years [±8.5 SD]) and 673 AA controls (688 women and 165 men, mean AAE 73.1 years [±5.4 SD]). As with the NHW cohort, all AA individuals enrolled self-identified as AA and this ethnicity was confirmed using high density SNP data and analysis [4,5].

2.2. Targeted sequencing of eight LOAD risk loci

Targeted sequencing was performed in 291 cases and 103 controls and targeted all genomic sequence of *ABCA7* (exonic, intronic and intergenic regions). Targeted sequencing used hybridization-based targeted capture with NimbleGen SeqCap EZ probe libraries and Illumina HiSeq 2000 sequencing [16]. The sequencing data were processed using the Illumina Real Time Analysis (RTA) base calling pipeline version 1.8. The Burrows-Wheeler Aligner (BWA) [17] was used to map sequences to the hg19 human reference genome and variant calling was performed with the Genome Analysis Toolkit (GATK) [18].

2.3. Variant filtering and selection of previously associated variants for follow-up genotyping

2.3.1. Variant filtering and follow-up genotyping of rare damaging variants overrepresented in cases

Variant quality control (QC) included removing variants with GQ <30, depth <8, VQSLOD <2, and call rate <90%. Variants passing QC were annotated using SeattleSeq [19] and prioritized for follow-up genotyping in independent case-control datasets using several criteria including: (1) functionality (missense, nonsense, or splice-site variant), (2) potentially damaging effect as defined by Polyphen-2[20], and (3) overrepresentation in cases. Overrepresentation in cases was defined by the variant meeting one or more criteria: 1) case variants absent in controls and Exome Variant Server (EVS) [21] and dbSNP [22], or 2) case-only variants having a case MAF ratio two times greater than the EVS. Variants identified as rare, damaging, and overrepresented in cases vs. controls were genotyped using the Sequenom Array in 4076 NHW familial and sporadic cases (N = 1987) and controls (N = 2089) and 1157 AA cases (N = 484) and controls (N = 673). Fig. 1 presents a summary flowchart of this strategy. This strategy prioritizes variants most likely to influence expression of a trait, but necessarily eliminates potential regulatory and protective variants.

2.3.2. Selection and genotyping of previously associated variants in the eight LOAD loci from other studies

Based on our knowledge of the literature, we selected several rare or low-frequency variants that have recently been associated with *ABCA7* [6–8] for follow-up 'confirmatory' genotyping and generalizability in 4076 NHW and 1157 AA cases and controls from the datasets described above. A total of nine variants were chosen and genotyped using Taqman assays (Supplementary Table 2).

2.4. Association testing

Association testing of variants and genes was performed in the discovery cohort. Permutation testing for association was performed in PLINK for individual variants [23]. Combined effects

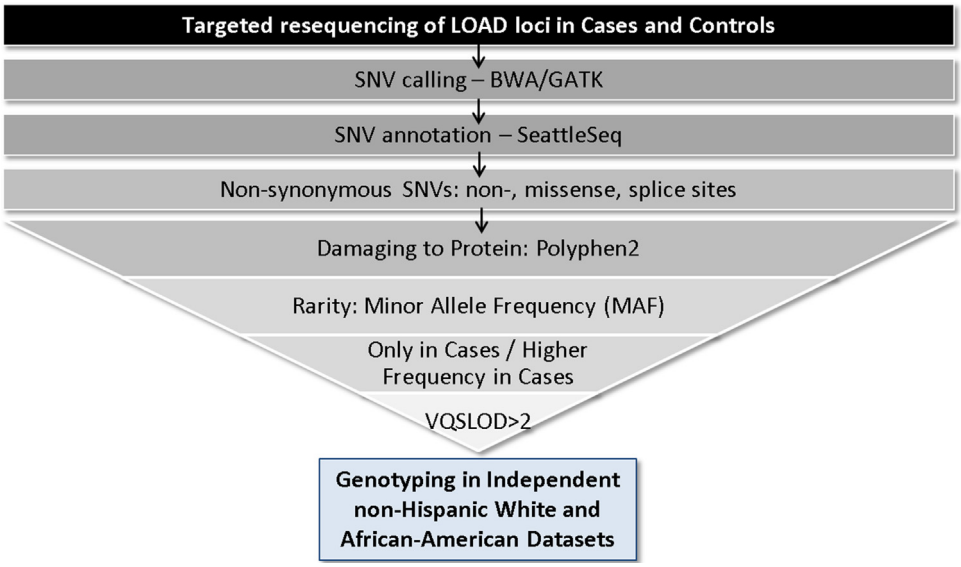


Fig. 1. Flowchart of variant filtering and follow-up genotyping in independent datasets.

of variants with MAF ≤ 0.05 in gene overall and regions of the gene was estimated using SKAT [24]. All analyses adjusted for age, sex and principal components. Principal components were derived from genotyping chip markers available on all individuals. Gene association testing assessed case vs. control differences in the effect of rare variants per genomic consequence (consequences: a. coding-missense/nonsense/splice; or b. synonymous coding, utr, intronic).

The GENMOD program, which corrects for familial relatedness and is part of the SAS/STAT software, and Fisher's exact test (for small sample sizes of cases and controls only) were used to analyze variants from the follow-up genotyping variants in the NHW and AA case-control replication dataset (Supplementary Table 1). Assessment of familial segregation of variants was conducted for the results in the NIMH sample. Results of these analyses were assessed further using allele frequencies from the Exome Aggregation Consortium (ExAC) [25], which were combined with the HIHG/CWRU control sample to increase sample size for testing of rare alleles [25]. AA data were combined with the African ExAC population (N = 5203) (ExAC AFR), and NHW results were combined with the non-Finnish European ExAC population (N = 33,370) (ExAC NFE) for the analyses. Due to the use of allele frequencies only (as individual genotypes and covariates are unavailable for ExAC), adjustment for covariates was not possible in the ExAC analyses. For genotyping of previously reported variants, the GENMOD program was utilized to perform the follow-up association tests under a generalized linear regression model. Association tests were performed with adjustments for age, sex, *APOE* status, and relatedness between samples. Follow-up of these analyses also utilized both NHW and AA ExAC allele frequencies in combination with our genotyped control allele frequencies. Conditional analysis of the rs7811724 variant was also performed with GENMOD, correcting for age, sex and PCs. Linkage disequilibrium (LD) between variants was assessed using Plink [23].

3. Results

3.1. Targeted sequencing

The 291 cases and 103 controls identified for sequencing have an average AAO for cases of 72.4 years and AAE for controls of 83.6 years (Table 1). 91% of the target region coverage was at a read depth of >80 over all samples.

3.2. Variant filtering

Filtering the targeted sequencing data on genotype quality (GQ) ≥ 30, depth ≥ 8, variant call rate ≥ 90%, and VQSLOD ≥ 2 produced a total of 233 high quality variants. Of these variants, 209 (89.7%) had MAF <5% and 138 (58.9%) were novel when compared to dbSNP. 14 rare variants were predicted to be possibly or probably damaging by Polyphen-2 and were present in one or more cases while absent in controls (average MAF = 0.002) (Supplementary Table 3). Eight of these variants and one splice variant in *ABCA7* were selected for follow-up genotyping in the two previously described NHW and AA case-control datasets (see Supplementary Table 4 for variant list).

3.3. SNV and gene/region-based association testing of targeted sequencing data from discovery dataset

Permutation testing of all 209 high-quality rare targeted sequencing variants identified one variant at MAF <0.05 with a nominally significant *p*-value ($P < 0.05$) of 0.003, an intronic variant present in controls only (MAF = 0.007) that does not survive Bonferroni correction for the number of variants tested. The small sample size of the discovery dataset is not ideal for single-variant testing, and therefore aggregate association tests were also conducted. However, gene-based association testing with SKAT of low-frequency (MAF ≤ 0.05) and rare (MAF ≤ 0.01) variants grouped by consequence found no clusters of variants associated with LOAD.

3.4. Follow-up genotyping association analyses

Follow-up genotyping of eight rare variants prioritized from the targeted sequencing identified three variants present in at least two cases or controls, which we assessed with association testing (Table 2). These included a 3-UTR splice variant in *ABCA7* (rs376824416) present only in NHW affecteds (N = 5), four of which are affected siblings of one family (see Fig. 2 for pedigree). Testing with NHW ExAC data (variant MAF = 0.00012) showed the variant to be nominally associated with LOAD ($P = 0.040$). The remaining five variants were either monomorphic in the follow-up genotyping dataset (N = 1) or were only seen in one case (N = 4) (Supplementary Table 5).

Table 2
Association testing results for follow-up genotyping variants.

				Heterozygous Allele Count (Affected/Unaffected)				ExAC Minor Allele Frequency[b]		Fisher's Exact P	
Chr:Position	Ref	Alt	Gene	All NHW	NHW	NIMH	AA	NFE	AA	NHW	AA
19:1041971	T	G	*ABCA7*	1/2	1/2	0/0	0/0	1.10E-03	1.30E-04	3.92E-01	–
19:1043238	G	T	*ABCA7*	1/1[a]	0/0	1/1	0/2	3.00E-04	1.90E-04	1.24E-01	1.00E+00
19:1054190	A	G	*ABCA7*	2/0[a]	1/0	4/0	0/0	1.50E-04	1.00E-04	4.08E−02*	–

[a] Counts 1 case/control per family.
[b] ExAC MAF from The Exome Aggregation Consortium [25].
* Nominally significant and fully segregating in family.

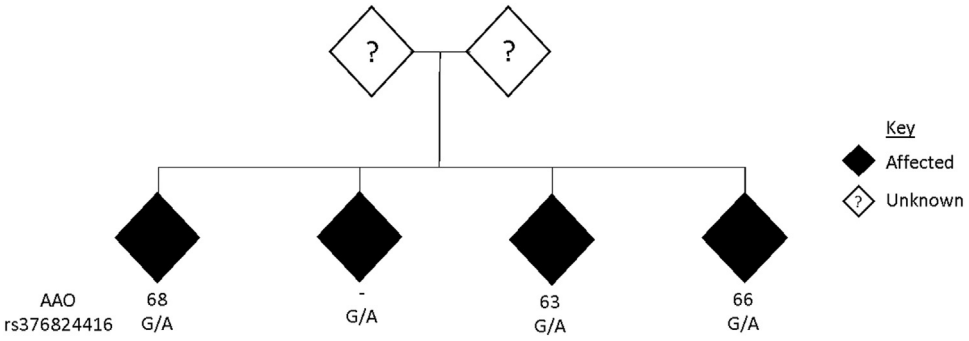


Fig. 2. Pedigree for NIMH family 380 showing full segregation of the *ABCA7* rs376824416 3-UTR splice variant. The age beneath each individual is the age at onset for individuals with LOAD (AAO). The genotypes are the wild type (G/G) or the alteration (G/A) that causes the *ABCA7* splicing alteration (rs376824416).

3.5. Follow-up of known rare and low-frequency variant associations in LOAD risk loci

Follow-up genotyping of previously reported rare and low-frequency variants in known AD loci confirmed the association of an intronic SNP (rs78117248) in the *ABCA7* gene in NHW Americans ($P = 3.0 \times 10^{-3}$), a result originally reported as associated with LOAD risk in a Belgian population [7]. Principal component analyses of 1000G European ethnic groups and the rs78117248 variant in HIHG samples showed the rs78117248 variant clustering with the British in England and Scotland (GBR) population, with a few carriers from the Iberian Population in Spain (IBS) population (Supplementary Fig. 1). rs78117248 is in complete LD (D' = 1.00) with the top associated *ABCA7* SNP (rs3764650) from the Hollingsworth et al. LOAD GWAS, though the r-squared is modest at 0.23. Association analysis of rs3764650 in our dataset yields $P = 0.054$, but this association is non-significant ($P = 0.380$) when we condition on the suspected risk variant rs78117248, a result similar to the Belgian cohort analyses which nominated this SNP as a risk variant [7].

Two other known variants were significantly associated with LOAD status, including a nominal association for the *ABCA7* variant rs200538373 (19:1061892:G > C) from Steinberg et al. [8] in NHW (OR = 2.12, $P = 0.027$), which also increased risk in AA, though the result was not significant (OR = 2.42, $P = 0.290$) (see Table 3 for results of all 10 variants). Association testing of these data using ExAC found significant associations with both NHW ($P = 1.2 \times 10^{-6}$) and AA ($P = 0.017$), though the AA association would not survive Bonferroni correction. This variant is not in LD with the previously identified risk deletion in ABCA7 ($r^2 = 0$), and thus potentially represents an independent risk factor for LOAD in AA. Additionally, we confirmed a previously reported NHW association of an *ABCA7* stop-gain variant (19:1058154:G > T) from Vardarajan et al. [6] ($P = 7.0 \times 10^{-3}$ in the follow-up genotyping dataset) and in combined testing with ExAC ($P = 1.0 \times 10^{-3}$).

4. Discussion

This work identified several rare variants in *ABCA7* as associated with risk of LOAD (see Fig. 3 for the genic location of these variants), including three loss-of-function (LOF) variants (two splice and one stop-gain variant) and an intronic variant. One of these associations is novel, a 3′-UTR splice acceptor variant (rs376824416) present only in affecteds (N = 5), including all 4 affected siblings of one family, that is nominally associated with LOAD when compared to ExAC frequencies. Interestingly, deletions [8], intronic variants [7], stop-gain variants [6], and splice donor variants [8] have been associated with LOAD status, so this finding extends the types of *ABCA7* variants that may be pathogenic for LOAD.

Follow-up genotyping also supported the *ABCA7* splice donor variant rs200538373, previously reported as associated with LOAD in an Icelandic population[8], as a risk variant in NHW and suggested it as a risk variant for AA populations. Testing with ExAC frequencies found the variant associated with LOAD in both populations. We had previously identified a deletion (rs142076058) in *ABCA7* that is significantly associated with AD risk in AA [5]. The additional association of *ABCA7* splice donor variant rs200538373 (19:1061892:G>C) in these data, though nominal, suggests that as is the case in the NHW population, it is likely that multiple *ABCA7* pathogenic variants also exist for the AA population. Importantly, *ABCA7* is the major risk factor for LOAD in AA [4], and this is to our knowledge, the first report of an *ABCA7* risk variant that is associated in both NHW and AA populations. It is also important to note that the approach of using ExAC frequencies for testing of associations does not allow for adjustment for covariates, and thus these associations should be interpreted with caution and warrant further validation in other cohorts.

We also confirmed an association of a low-frequency intronic SNP (rs78117248) in NHW individuals from the USA, a result originally reported in a Belgian population [7]. This SNP is in LD (D' = 1.0, r2 = 0.23) with one of the *ABCA7* GWAS association variants [3], and conditional analysis supported the variant as a driver of the GWAS

Table 3
Results of genotyping of known associated variants.

Chr:Position*	Ref	Alt	ID	Gene	Follow-up Genotyping Results	
					NHW OR (95% CI); P-value	AA OR (95% CI); P-value
19:1047176	C	A/G	–	*ABCA7*	Variant not observed	Variant not observed
19:1047507	GGAGCAG	–	–	*ABCA7*	**	**
19:1052853	A	G	rs78117248	*ABCA7*	1.56 (1.16–2.11); 0.003*	0.75 (0.27–2.07); 0.58
19:1056244	T	G	rs113809142	*ABCA7*	2.19 (0.18–26.03); 0.53	Variant not observed
19:1058154	G	T	–	*ABCA7*	6.55 (1.66–25.75); 0.007*	Variant not observed
19:1061892	G	C	rs200538373	*ABCA7*	2.12 (1.08–1.89); 0.027	2.42 (0.46–12.73); 0.29

Positions are in NCBI 37.
* Significant in ExAC testing after Bonferroni correction for 5 variants tested.
** Failed genotyping.

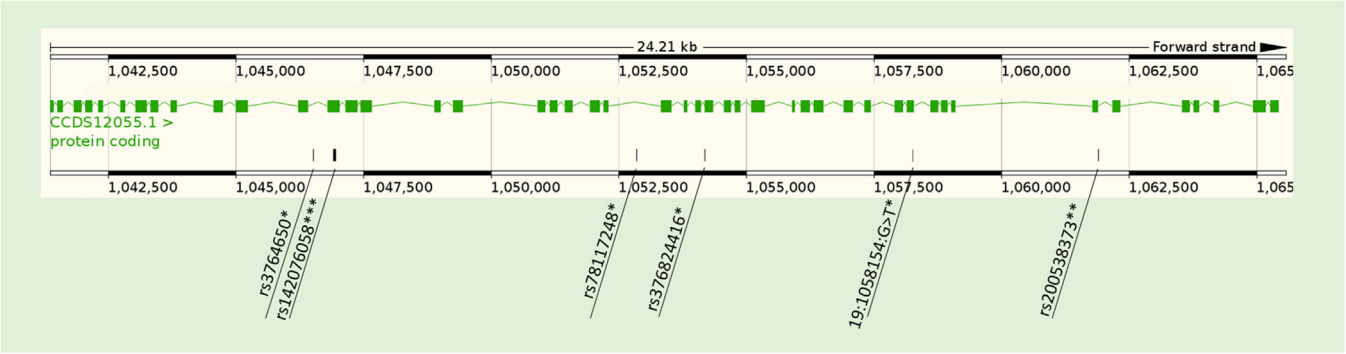


Fig. 3. Genomic location of ABCA7 risk associated variants. *Non-Hispanic White risk variant, **non-Hispanic White and African-America risk variant, ***African-American and Caribbean Hispanic risk variant.

signal, suggesting its potential as a causal variant for the locus. As noted in Cuyvers et al., 2015 [7], this SNP is located in a transcription factor binding region and has regulatory potential according to RegulomeDB [26]. According to 1000 genomes populations MAFs [27] the variant is not present in Africans or East Asians, but is present in Europeans (MAFs = 3% in Italians, 2.5% in Caucasians from Utah, 2% in British and Spanish, and 1% in Finnish), Admixed Americans (MAF = 2.3% with sub-population MAFs of 4% in Puerto Ricans, 2% in Mexicans and Colombians, and 1% in Peruvians), and South Asians (MAF = 0.5%), suggesting the variant could have been propagated from European populations during colonial periods.

Finally, we confirmed a stop-gain variant (19:1058154:G>T), originally associated in an NHW population, as significantly associated with LOAD risk in our NHW population. Thus, we identified several rare, functional variants in *ABCA7* associated with risk of LOAD, suggesting that multiple rare variants in this gene contribute to LOAD risk. Importantly, our data show this is likely the case for both AA and NHW populations. These and other previously reported variants in *ABCA7* have differing biological consequences, but many appear to lead to LOF of *ABCA7*[6–8], including two splicing variants and a stop-gain variant we associate with LOAD in this work. *ABCA7* has been associated with both neuritic plaque pathology [28] and neurofibrillary tangle pathology [29] in LOAD, and while the mechanisms by which *ABCA7* confers risk of LOAD are still being determined, functional studies are starting to support LOF as a pathological mechanism leading to LOAD [30,31]. Notably, SPIDEX [30] assigns the splicing variants identified here, rs200538373 and rs376824416, significant values for increased exon skipping (percent of transcripts with the central exon spliced in z-scores equal to −3.500 and −2.552, respectively), indicating that these variants are likely to lead to decreased expression or a LOF of *ABCA7*. Furthermore, research over the past decade or more has revealed the potential importance of splicing in the pathogenesis of both LOAD [32,33] and other neurodegenerative diseases [34–36], providing mechanistic support to splicing variants involvement in LOAD. In addition, among LOAD GWAS loci, *ABCA7* has been shown to have the largest number of splicing changes in LOAD cases compared to controls [33]. Further confirmation and characterization of these, and other *ABCA7* LOAD risk variants, will be important for development of therapeutic strategies targeting *ABCA7*.

Acknowledgements

We are grateful to the many patients with Alzheimer's disease and their families for their participation, without whom this work would not be possible. This research was supported by grants from the National Institutes of Health (1R01AG027944, UO1AG032984, U01AG052410), the Alzheimer's Association (IIRG09133827), and the BrightFocus Foundation (A2011048).

NIHM Study Samples: Data and biomaterials were collected in three projects that participated in the National Institute of Mental Health (NIMH) Alzheimer Disease Genetics Initiative. From 1991-98, the Principal Investigators and Co-Investigators were: Massachusetts General Hospital, Boston, MA, U01 MH46281, Marilyn S. Albert, Ph.D., and Deborah Blacker, M.D., Sc.D.; Johns Hopkins University, Baltimore, MD, U01 MH46290, Susan S. Bassett, Ph.D., Gary A. Chase, Ph.D., and Marshal F. Folstein, M.D.; University of Alabama, Birmingham, AL, U01 MH46373, Rodney C.P. Go, Ph.D., and Lindy E. Harrell, M.D.

NCRAD Study Samples: Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neulet.2017.04.014.

References

[1] J.C. Lambert, C.A. Ibrahim-Verbaas, D. Harold, A.C. Naj, R. Sims, et al., Meta-analysis of 74, 046 individuals identifies 11 new susceptibility loci for Alzheimer's disease, Nat. Genet. 45 (2013) 1452–1458, http://dx.doi.org/10.1038/ng.2802.

[2] A.C. Naj, G. Jun, G.W. Beecham, L.S. Wang, B.N. Vardarajan, et al., Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease, Nat. Genet. 43 (2011) 436–441, http://dx.doi.org/10.1038/ng.801.

[3] P. Hollingworth, D. Harold, R. Sims, A. Gerrish, J.C. Lambert, et al., Common variants at ABCA7, MS EPHA1, CD33 and CD2AP are associated with Alzheimer's disease, Nat. Genet. 43 (2011) 429–435, http://dx.doi.org/10.1038/ng.803.

[4] C. Reitz, G. Jun, A. Naj, R. Rajbhandary, B.N. Vardarajan, et al., Alzheimer Disease Genetics, Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E 4, and the risk of late-onset Alzheimer disease in African Americans, JAMA 309 (2013) 1483–1492, http://dx.doi.org/10.1001/jama.2013.2973.

[5] H.N. Cukier, B.W. Kunkle, B.N. Vardarajan, S. Rolati, K.L. Hamilton-Nelson, M.A. Kohli, P.L. Whitehead, B.A. Dombroski, D. Van Booven, R. Lang, D.M. Dykxhoorn, L.M. Farrer, Mi.L. Cuccaro, J.M. Vance, J.R. Gilbert, G.W. Beecham, E.R. Martin, R.M. Carney, R. Mayeux, G.D. Schellenberg, G.S. Byrd, J.L. Haines, M.A. Pericak-Vance, ABCA7 Frameshift Deletion Associated with Alzheimer's Disease in African Americans, Neurol. Genet. 2 (3) (2016) e79.

[6] B.N. Vardarajan, M. Ghani, A. Kahn, S. Sheikh, C. Sato, S. Barral, J.H. Lee, R. Cheng, C. Reitz, R. Lantigua, D. Reyes-Dumeyer, M. Medrano, I.Z. Jimenez-Velazquez, E. Rogaeva, P. St George-Hyslop, R. Mayeux, Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci, Ann. Neurol. 78 (2015) 487–498, http://dx.doi.org/10.1002/ana.24466.

[7] E. Cuyvers, A. De Roeck, T. Van den Bossche, C. Van Cauwenberghe, K. Bettens, S. Vermeulen, M. Mattheijssens, K. Peeters, S. Engelborghs, M. Vandenbulcke, R. Vandenberghe, P.P. De Deyn, C. Van Broeckhoven, K. Sleegers, Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study, Lancet Neurol. 14 (2015) 814–822, http://dx.doi.org/10.1016/s1474-4422(15)00133-7.

[8] S. Steinberg, H. Stefansson, T. Jonsson, H. Johannsdottir, A. Ingason, H. Helgason, P. Sulem, O.T. Magnusson, S.a. Gudjonsson, U. Unnsteinsdottir, A. Kong, S. Helisalmi, Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease, Nat. Genet. (2015) 26–29, http://dx.doi.org/10.1038/ng.3246.

[9] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, E.M. Stadlan, Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease, Neurology 34 (1984) 939–944 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=6610841.

[10] G.M. McKhann, D.S. Knopman, H. Chertkow, B.T. Hyman, C.R. Jack, C.H. Kawas, W.E. Klunk, W.J. Koroshetz, J.J. Manly, R. Mayeux, R.C. Mohs, J.C. Morris, M.N. Rossor, P. Scheltens, M.C. Carrillo, B. Thies, S. Weintraub, C.H. Phelps, 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. 7 (2011) 263–269, http://dx.doi.org/10.1016/j.jalz.2011.03.005.

[11] G.W. Beecham, E.R. Martin, Y.J. Li, M.A. Slifer, J.R. Gilbert, J.L. Haines, M.A. Pericak-Vance, Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease, Am. J. Hum. Genet. 84 (2009) 35–43, http://dx.doi.org/10.1016/j.ajhg.2008.12.008.

[12] G. Jun, A.C. Naj, G.W. Beecham, L.S. Wang, J. Buros, et al., Meta-analysis confirms CR1, CLU, and PICALM as alzheimer disease risk loci and reveals interactions with APOE genotypes, Arch. Neurol. 67 (2010) 1473–1484, http://dx.doi.org/10.1001/archneurol.2010.201.

[13] The National Institute of Mental Health (NIMH) Genetic Studies of Alzheimer's Disease Cohort, (2016). https://www.nimhgenetics.org/available_data/alzheimers_disease/.

[14] D. Blacker, J.L. Haines, L. Rodes, H. Terwedow, R.C. Go, L.E. Harrell, R.T. Perry, S.S. Bassett, G. Chase, D. Meyers, M.S. Albert, R. Tanzi, ApoE-4 and age at onset of Alzheimer's disease: the NIMH genetics initiative, Neurology 48 (1997) 139–147 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9008509.

[15] J.H. Lee, R. Cheng, N. Graff-Radford, T. Foroud, R. Mayeux, G. National Institute on Aging Late-Onset Alzheimer's Disease Family Study, Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci, Arch. Neurol. 65 (2008) 1518–1526, http://dx.doi.org/10.1001/archneur.65.11.1518.

[16] M.A. DePristo, E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, A.A. Philippakis, G. del Angel, M.A. Rivas, M. Hanna, A. McKenna, T.J. Fennell, A.M. Kernytsky, A.Y. Sivachenko, K. Cibulskis, S.B. Gabriel, D. Altshuler, M.J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data, Nat. Genet. 43 (2011) 491–498, http://dx.doi.org/10.1038/ng.806.

[17] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (2009) 1754–1760, http://dx.doi.org/10.1093/bioinformatics/btp324.

[18] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. a DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (2010) 1297–1303, http://dx.doi.org/10.1101/gr.107524.110.

[19] SeattleSeq, (2016). http://snp.gs.washington.edu/SeattleSeqAnnotation138/.

[20] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, S.R. Sunyaev, A method and server for predicting damaging missense mutations, Nat. Methods 7 (2010) 248–249, http://dx.doi.org/10.1038/nmeth0410-248.

[21] Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/), (2016).

[22] E.M. Smigielski, K. Sirotkin, M. Ward, S.T. Sherry, dbSNP: a database of single nucleotide polymorphisms, Nucleic Acids Res. 28 (2000) 352–355 http://www.ncbi.nlm.nih.gov/pubmed/10592272.

[23] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly, P.C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (2007) 559–575, http://dx.doi.org/10.1086/519795.

[24] M.C. Wu, S. Lee, T. Cai, Y. Li, M. Boehnke, X. Lin, Rare-variant association testing for sequencing data with the sequence kernel association test, Am. J. Hum. Genet. 89 (2011) 82–93, http://dx.doi.org/10.1016/j.ajhg.2011.05.029.

[25] Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org). (2016).

[26] A.P. Boyle, E.L. Hong, M. Hariharan, Y. Cheng, M. a Schaub, M. Kasowski, K.J. Karczewski, J. Park, B.C. Hitz, S. Weng, J.M. Cherry, M. Snyder, Annotation of functional variation in personal genomes using RegulomeDB, Genome Res. 22 (2012) 1790–1797, http://dx.doi.org/10.1101/gr.137323.112.

[27] A. Auton, G.R. Abecasis, D.M. Altshuler, R.M. Durbin, G.R. Abecasis, et al., A global reference for human genetic variation, Nature 526 (2015) 68–74, http://dx.doi.org/10.1038/nature15393.

[28] J.M. Shulman, K. Chen, B.T. Keenan, L.B. Chibnik, A. Fleisher, P. Thiyyagura, A. Roontiva, C. McCabe, N.A. Patsopoulos, J.J. Corneveaux, L. Yu, M.J. Huentelman, D.A. Evans, J.A. Schneider, E.M. Reiman, P.L. De Jager, D.A. Bennett, Genetic susceptibility for Alzheimer disease neuritic plaque pathology, JAMA Neurol. 1–7 (2013), http://dx.doi.org/10.1001/jamaneurol.2013.2815.

[29] L. Yu, L.B. Chibnik, G.P. Srivastava, N. Pochet, J. Yang, J. Xu, J. Kozubek, N. Obholzer, S.E. Leurgans, J. a Schneider, A. Meissner, P.L. De Jager, D. a Bennett, Association of brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of alzheimer disease, JAMA Neurol. 60612 (2014) 1–10, http://dx.doi.org/10.1001/jamaneurol.2014.3049.

[30] N. Sakae, C.-C. Liu, M. Shinohara, J. Frisch-Daiello, L. Ma, Y. Yamazaki, M. Tachibana, L. Younkin, A. Kurti, M.M. Carrasquillo, F. Zou, D. Sevlever, G. Bisceglio, M. Gan, R. Fol, P. Knight, M. Wang, X. Han, J.D. Fryer, M.L. Fitzgerald, Y. Ohyagi, S.G. Younkin, G. Bu, T. Kanekiyo, ABCA7 deficiency accelerates amyloid- generation and alzheimer's neuronal pathology, J. Neurosci. 36 (2016) 3848–3859, http://dx.doi.org/10.1523/JNEUROSCI.3757-15.2016.

[31] K. Satoh, S. Abe-Dohmae, S. Yokoyama, P. St. George-Hyslop, P.E. Fraser, ATP-binding cassette transporter A7 (ABCA7) loss of function alters Alzheimer amyloid processing, J. Biol. Chem. 290 (2015) 24152–24165, http://dx.doi.org/10.1074/jbc.M115.655076.

[32] J.R. Tollervey, Z. Wang, T. Hortobagyi, J.T. Witten, K. Zarnack, M. Kayikci, T.A. Clark, A.C. Schweitzer, G. Rot, T. Curk, B. Zupan, B. Rogelj, C.E. Shaw, J. Ule, Analysis of alternative splicing associated with aging and neurodegeneration in the human brain, Genome Res. 21 (2011) 1572–1582, http://dx.doi.org/10.1101/gr.122226.111.

[33] C. Humphries, M.A. Kohli, P. Whitehead, D.C. Mash, M.A. Pericak-Vance, J. Gilbert, Alzheimer disease (AD) specific transcription, DNA methylation and splicing in twenty AD associated loci, Mol. Cell. Neurosci. 67 (2015) 37–45, http://dx.doi.org/10.1016/j.mcn.2015.05.003.

[34] F. Liu, C.-X. Gong, Tau exon 10 alternative splicing and tauopathies, Mol. Neurodegener. 3 (2008) 8, http://dx.doi.org/10.1186/1750-1326-3-8.

[35] J.D. Mills, M. Janitz, Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases, Neurobiol. Aging 33 (2012) 1012, http://dx.doi.org/10.1016/j.neurobiolaging.2011.10.030, e11-1012. e24.

[36] K. Beyer, A. Ariza, Alpha-synuclein posttranslational modification and alternative splicing as a trigger for neurodegeneration, Mol. Neurobiol. (2012) 1–16, http://dx.doi.org/10.1007/s12035-012-8330-5.

HHS Public Access

Author manuscript

J Alzheimers Dis Parkinsonism. Author manuscript; available in PMC 2017 November 22.

Published in final edited form as:

J Alzheimers Dis Parkinsonism. 2017 August ; 7(4): . doi:10.4172/2161-0460.1000355.

Exome Sequencing of Extended Families with Alzheimer's Disease Identifies Novel Genes Implicated in Cell Immunity and Neuronal Function

HN Cukier[1,2,†], **BK Kunkle**[1,†], **KL Hamilton**[1], **S Rolati**[1], **MA Kohli**[1], **PL Whitehead**[1], **J Jaworski**[1], **JM Vance**[1,3], **ML Cuccaro**[1,3], **RM Carney**[1,4], **JR Gilbert**[1,3], **LA Farrer**[5], **ER Martin**[1,3], **GW Beecham**[1,3], **JL Haines**[6], and **MA Pericak-Vance**[1,2,3,*]

[1]John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA

[2]Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA

[3]John T. Macdonald Foundation, Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA

[4]Mental Health and Behavioral Sciences Service, Miami Veterans Affairs, Miami, FL, USA

[5]Departments of Medicine, Neurology, Ophthalmology, Genetics and Genomics, Epidemiology and Biostatistics, Boston University, Boston, MA, USA

[6]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA

Abstract

Objective—Alzheimer's disease (AD) is a neurodegenerative disorder for which more than 20 genetic loci have been implicated to date. However, studies demonstrate not all genetic factors have been identified. Therefore, in this study we seek to identify additional rare variants and novel genes potentially contributing to AD.

Methods—Whole exome sequencing was performed on 23 multi-generational families with an average of eight affected subjects. Exome sequencing was filtered for rare, nonsynonymous and loss-of-function variants. Alterations predicted to have a functional consequence and located within either a previously reported AD gene, a linkage peak (LOD>2), or clustering in the same gene across multiple families, were prioritized.

Results—Rare variants were found in known AD risk genes including *AKAP9, CD33, CR1, EPHA1, INPP5D, NME8, PSEN1, SORL1, TREM2* and *UNC5C*. Three families had five variants of interest in linkage regions with LOD>2. Genes with segregating alterations in these peaks include *CD163L1* and *CLECL1*, two genes that have both been implicated in immunity, *CTNNA1*,

[*]**Corresponding author:** Pericak-Vance MA, Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL USA, Tel: 3052432308; Fax: 3052432396; mpericak@med.miami.edu.

[†]Authors contributed equally

which encodes a catenin in the cerebral cortex and *MIEF1*, a gene that may induce mitochondrial dysfunction and has the potential to damage neurons. Four genes were identified with alterations in more than one family include *PLEKHG5*, a gene that causes Charcot-Marie-Tooth disease and *THBS2*, which promotes synaptogenesis.

Conclusion—Utilizing large families with a heavy burden of disease allowed for the identification of rare variants co-segregating with disease. Variants were identified in both known AD risk genes and in novel genes.

Keywords

Alzheimer's disease; dominant inheritance; linkage; multiplex; whole exome sequencing

Introduction

Alzheimer's disease (AD) is the leading cause of dementia in the elderly [1]. The majority of individuals present with late-onset AD (≥ 65 years), but early-onset (<65 years) has also been reported in ~5% of cases. Both common genetic variants, such as the *APOE ε4* allele, and rare variants, have been found to impact the risk for both early- and late-onset AD [2–5]. While more than 20 genetic loci have been connected with late-onset AD to date, the underlying genetic architecture is complex and new risk genes are still being identified [6].

While genome-wide association studies (GWAS) have been key in identifying a majority of the novel regions of genetic risk in the past ten years, by design, GWAS are unlikely to recognize risk variants with rare frequencies in the population and necessitate the use of large cohorts of hundreds or even thousands of individuals to reach statistically significant conclusions [6]. In contrast, whole exome sequencing (WES) provides an alternative and complementary method to locate rare alterations in genes which may have medium to large effects on disease risk and require far fewer participants [6–8]. WES studies have identified new mutations in both known AD genes and novel risk genes, including *AKAP9, PLD3, TREM2* and *UNC5C*, as well as protective variants, such as those in *TREML2* [7–17]. Moreover, studying families with a heavy burden of AD and searching for genetic changes that segregate with disease can provide a unique opportunity to locate rare variants in novel risk genes such as *NOTCH3, PLD3* and *TTC3* [9,13,18]. These large AD families can reveal how multiple genetic variants may act in concert to influence risk [19–21]. For example, the *APOE ε2* allele was found to delay the age of onset by ~12 years in carriers of the E280A mutation in the *PSEN1* gene in the early-onset 'Paisa' pedigree [19]. In addition, genetic linkage can assist in narrowing genomic regions of interest potentially related to disease in large families [22]. In an effort to discover novel genes that may contribute to late-onset AD risk, we performed WES in 23 multiplex families that present with dominant inheritance patterns and prioritized variants that were inherited from common ancestors.

Materials and Methods

Patient ascertainment of extended AD families

240 individuals (77 AD subjects, 4 individuals with mild cognitive impairment (MCI) and 159 unaffected relatives) from 23 families of European ancestry heavily affected with late-

onset AD were utilized in this study (Supplementary Table 1). All family members were recruited after providing informed consent and with approval by the relevant institutional review boards. Affected individuals meet the standard NINCDS-ADRDA criteria for AD and MCI [23–25]. In addition, cognitive and neuropsychiatric data were collected on all affected indivduals using the NCRAC LOAD battery, the Geriatric Depression Scale (GDS15), the Cornell Scale for Depression in Dementia (CSDD) and the Neuropsychiatric Inventory Questionnaire (NPIQ).

Whole exome sequencing and variant detection

99 individuals (77 AD patients, 4 individuals with MCI, and 18 unaffected relatives) from 23 AD extended families underwent WES (Supplementary Table 1). Three micrograms of DNA from each sample were prepared using the SureSelect Human All Exon 50Mb Kit (Agilent Technologies) and the Paired-End Multiplexed Sequencing library kit (Illumina). Exome capture and sequence library construction was performed on a Sciclone G3 NGS Workstation (Caliper Life Sciences) and DNA was tested for uniform enrichment of targets with qPCR following established protocols provided by Agilent. Two exome sample libraries were sequenced per lane on a HiSeq 2000 Sequencing System (Illumina) in paired-end 2 × 100 bp runs. Sequencing data was processed using the Illumina RTA base calling pipeline v1.8. Reads were aligned to the human reference genome (hg19) with the Burrows-Wheeler Aligner (BWA) and variant calling performed with the Genome Analysis Toolkit (GATK) version 2.8 [26,27]. GATK parameters for variant quality control included duplicate sequence read removal, minimum read depth of 5, genotype quality (GQ) ≥ 20, variant quality score recalibration (VQSR, VQSLOD>0) and Genome Mappability Scores equal to 1 for the 35 base pair (bp) track and greater than or equal to 0.5 for the 20 bp track from the Duke Uniqueness Track [28]. The Duke uniqueness scores, generated for the ENCODE project and available as tracks in the University of California, Santa Cruz (UCSC) Genome Browser, report how unique a sequence is, where scores of 1 represent a completely unique sequence, a score of 0.5 indicates the sequence occurs exactly twice, and 0 represents the sequence occurs >4 times in the genome [29,30]. Small insertions and deletions were recognized by aligning the data with Bowtie2 and analyzing with the Pindel program [31,32].

Genotyping and variant filtering

234 individuals, including all 99 samples that had WES, were evaluated by genome-wide SNP (single nucleotide polymorphism) arrays including the Human 1Mv1 BeadChip, the 1M-DuoV3 BeadChip, the HumanOmniExpress-12 v1.0 BeadChip, and the HumanOmni2.5-4v1 BeadChip. All chips were processed using the Tecan EVO-1 robot and BeadChips were scanned with either the Illumina BeadArray Reader or iScan. Data was extracted by the Genome Studio software and a GenCall cutoff score of 0.15 was used. Samples were required to have a genotyping call rate of 98% or higher, and SNPs a call rate of 95% or greater, to pass quality control. SNPs were only included in the analysis if they were present in at least 60% of samples across all platforms. Checks for relatedness, Mendelian inconsistencies, gender based on X-chromosome heterozygosity, and concordance between the genotypes of the variants identified through exome sequencing and

genotyping were evaluated with PLINK version 1.07 [33]. All samples passed the quality control metrics.

Genotyping information was further used to delineate identical by descent (IBD) regions within each multiplex AD family. IBD filtering was implemented through the extended haplotype procedure in MERLIN version 1.1.2 [34]. Regions shared across all available AD individuals within a family were used to determine the IBD sharing segments and were, therefore, unique within each family. To determine the start and stop positions of IBD sharing regions within each family, the MERLIN output was evaluated in a sliding window of ten SNPs, defining IBD as sharing at each location with a threshold >50%.

Linkage analysis

Nonparametric and parametric two-point and multipoint linkage analyses were performed using MERLIN [32]. A disease allele frequency of 0.0001 was used in an affecteds-only model for parametric analysis. PLINK was employed for LD pruning in the multipoint analysis, with CEU HapMap data as the reference population and the following settings: the indep-pairwise option with a window size of 50, a step of 5 and an r^2 threshold of 0.5 [33,34].

Variant annotation and prioritization

Alterations passing quality measurements were annotated with the KGGSeq and ANNOVAR programs [35,36]. Variants were normalized prior to annotation [37]. Ensembl, RefSeq, and Gencode transcripts were all annotated, and the top consequence per gene was used for prioritization. CADD v1.3 scores were downloaded from the CADD server (http://cadd.gs.washington.edu/home) [38]. Figure 1 is an overview of the filtering and prioritization strategies used in this study. Brief descriptions of our three prioritization strategies are described below.

Variants in reported AD genes or loci—For all of the families, we evaluated whether variants were located in known AD risk genes; this includes genes identified in both early (*APP, PSEN1, PSEN2, GRN* and *TREM2*) and late-onset AD (Supplemental Table 2) [3,4]. Variants of interest were restricted to those with a minor allele frequency (MAF) ≤ 2% in the Kaviar Genomic Variant Database (version 160204-Public, 77,781 individuals) since these genes are known loci for AD [39]. The top variants of interest were validated by traditional Sanger sequencing.

Families with LOD scores >2—For each of the families, variants that segregated in all sequenced, affected individuals within areas LOD>2 were evaluated. Variants with a global MAF ≤ 1% in the Kaviar Database were prioritized. A MAF cutoff of ≤ 1% was implemented because variants with a MAF>1% in any ethnic population are unlikely to be a highly penetrant risk variant for AD [39–41]. This stricter MAF criteria was utilized to attempt to identify novel risk genes as opposed to variants in known AD genes. Variants were also prioritized based on their potential pathogenicity with the Combined Annotation-Dependent Depletion (CADD) score; scores ≥ 15 are predicted to be more likely to

contribute to a disease risk as this score represents “the median value for all possible canonical splice site changes and non-synonymous variants” [38].

Variants and genes shared across families—Analysis across all 23 families was performed to identify if there were any genes with rare, nonsynonymous or loss-of-function (LOF) variants in more than one family. Variants were selected that had a MAF ≤ 1% in the Kaviar Database and CADD scores ≥ 15 to try to identify potentially damaging alterations [39].

Association testing of top candidates

All top variants and genes from the three separate analyses described above were evaluated as potential risk variants using genome-wide association statistics for two family study cohorts (NIA-LOAD and MIRAGE) in the Alzheimer Disease Genetics Consortium [42,43]. Both gene and SNP-based tests were adjusted for age, sex and principal components (PCs). SNP-based logistic regression tests in each study were performed in the SNPTest program, and meta-analysis of these results was conducted using METAL [21,44]. Gene-based tests were conducted on meta-analysis summary statistics using VEGAS [44]. Variants tested in the gene-based analysis included all variants with a MAF<5%.

Results

Variants identified in known AD genes

Each sequenced family contained between 4 and 16 individuals diagnosed with AD. Mean age-at-onset across all families was 74.3 years. We identified 14 potentially damaging variants in 10 known AD genes and GWAS implicated loci (Table 1). Seven of the variants were observed in multiple affected individuals in the same family, while the remaining variants were observed only once. All alterations were single nucleotide changes with the exception of a four base pair deletion in *CD33*. This deletion is potentially the most deleterious as it is predicted to causes a frameshift that encodes two incorrect amino acids before terminating prematurely, thus failing to generate over 40% of the protein. In addition, multiple variants were observed in four genes: *AKAP9, INPP5D, SORL1* and *UNC5C*. Each gene had at least two families with a variant identified in it, while family 191 have a single affected individual with two alterations in *UNC5*. One of the variants in *UNC5C*, Ala860Thr, was identified in two different families; this alteration has a CADD score of 33, the highest score in this category.

Segregating variants in linkage regions

Linkage scans aggregating all families identified one primary linkage region, a parametric multipoint peak on chromosome 1q23 (161.9–165.6 MB). Two families had strong linkage in this region (family specific LOD>2). However, no variants met our filtering criteria for these two families, suggesting the causal variant(s) may be non-coding changes either removed from by our filtering criteria or not present in our WES. Three of the 23 families also have family-specific parametric LOD scores >2; rare, potentially damaging alterations in five genes occurred within these regions and may potentially be the strongest novel AD

candidate genes (Table 2). The five alterations were all missense changes in *CD163L1, CLECL1, CTNNA1, GALR3* and *MIEF1*.

Genes with variants in more than one family

We identified four genes that had rare (MAF ≤ 1%), segregating, and potentially deleterious variants in at least two families (Table 3). Three of these genes had the same missense alteration identified in distinct families: *MKL2, PLEKHG5* and *THBS2*.

Association testing of variants and genes

From our prioritized sets, a total of 9 SNPs and 14 genes were available for testing in the ADGC family-based meta-analysis datasets (NIA-LOAD and Mirage). None of the variants tested were significantly associated with disease (Supplemental Table 3). The gene *MIEF1*, identified as a candidate gene in a family 1201 with rare, potentially damaging segregating variants in a region with a LOD score of 2.22, reached nominal significance (p=0.049, Supplemental Table 4).

Discussion

Through WES of large families with a heavy burden of AD, variants in both known and novel loci were identified that could contribute to risk. Filtering for rare, segregating, and potentially damaging variants identified five novel candidate genes (Table 2). These genes encompass a variety of functions that are suggestive of a link to AD. For example, two of these genes are involved in regulating immunity: *CD163L1* and *CLECL1* [45–47]. *CD163L1* is expressed in macrophages, upregulated in response to IL-10 and acts as an endocytic receptor [48]. *CLECL1* is highly expressed in B cells and dendritic cells and may enhance the immune response through upregulation of IL-4 [46]. Neuroinflammation has been shown to occur in AD patients, possibly through misregulation of microglia and triggered by amyloid beta plaques [49]. Additionally, established AD risk genes, such as *ABCA7, CD33* and *TREM2*, have also been linked to the immune system [4]. Another gene identified through this study, *CTNNA1*, encodes a catenin expressed at elevated levels in the nervous system [50]. *GALR3* is a receptor for the neuropeptide galanin, which has been shown to modulate a variety of processes, including cognition and memory, functions disrupted in AD [51,52]. *MIEF1* was nominally associated with late-onset AD in a meta-analysis of two family datasets from the ADGC, thereby suggesting that it may play a wider role in AD that extends beyond a single multiplex AD family. *MIEF1* may play a role in dysfunctional mitochondria and their potential to damage neurons [53,54]. Thus, each of the genes in the peak linkage regions are connected to known AD functions or neuronal pathways.

Four genes had rare, potentially damaging variants in more than one family (Table 3). When evaluating known functions of these genes, two are of particular interest, *PLEKHG5* and *THBS2*. *PLEKHG5* has been previously implicated in both Charcot-Marie-Tooth disease and spinal muscular atrophy [55,56]. *PLEKHG5* is ubiquitously expressed throughout the nervous system and murine studies demonstrated lowered expression can alter the velocity of nerve conduction [55,56]. In addition, *THBS2* is an intriguing novel AD candidate gene

due to its involvement in synaptogenesis in immature astrocytes [57]. Further investigation into each of these novel AD candidates and the variants identified in this study is required.

After evaluating our families for rare alterations in known AD genes and loci, variants were discovered in genes previously connected to both early and late-onset AD (Table 1). Four genes had multiple alterations: *AKAP9, INPP5D, SORL1* and *UNC5C*. Some of these alterations have the potential to interfere with a protein's function due to their location within specific domains. For example, a rare alteration in *UNC5C* identified in two distinct families, Ala860Thr, falls within the highly conserved DEATH domain, a region composed of alpha-helices and involved in apoptotic functions. Another study identified a different alteration in the same region in AD families and proposed that the alteration may increase the susceptibility of neurons to death [17]. A single affected individual in family 2349 was found to carry a frameshift deletion in *CD33* predicted to remove over 40% of the protein. There is evidence that higher expression of *CD33* in brains is associated with cognitive decline [58]. However, it may be that dysregulation of the protein, either through over or under expression, could contribute to AD risk. In addition, a rare alteration in *SORL1*, Thr588Ile, was identified within the vacuolar protein sorting 10 (VPS10) domain and may influence the processing of amyloid beta fragments, as has been shown for other AD associated variants in this gene [59]. Moreover, a potentially pathogenic alteration was identified in *PSEN1* in a single individual, Glu318Gly; this variant was previously reported to result in higher tau and phosphorylated tau levels in cerebrospinal fluids [60]. Three affected individuals from family 1893 were found to share the Arg336His alteration in the *NME8* gene, a change that fell within the first NDK domain of the protein. This gene has been associated with clinical features of AD including atrophy of the hippocampus and occipital gyrus [6,61]. These alterations, while not segregating within all affected individuals in the families, may play a contributing role in AD risk.

Conclusion

This study demonstrates how using large, extended families to evaluate exome data identifies segregating risk variants in potentially novel AD candidate genes. In contrast to GWAS studies that have grown from hundreds to thousands and tens of thousands of participants, this study design requires far fewer participants. Indeed, a single extended family may be sufficient to identify a novel AD candidate gene [18]. Moreover, WES has the sensitivity to directly detect both common and rare variants that may confer a risk to AD, while GWAS findings are limited to pinpointing a region of interest, but not necessarily the causative alterations. In the study presented here, rare changes potentially contributing to AD risk were found in genes implicated in the immune response, *CD163L1* and *CLECL1*, and neuronal function, *CTNNA1, GALR3, MIEF1, PLEKHG5* and *THBS2*. Variants were also identified in genes previously connected to both early and late-onset AD including *AKAP9, INPP5D, SORL1* and *UNC5C*. Further investigation will be required to fully assess the cellular and molecular consequences of the alterations identified here as well as determine whether the novel genes found are involved in AD risk across larger datasets.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (R01 AG027944, R01 AG028786 to MAPV, R01 AG019085 to JLH, P20 MD000546); a joint grant from the Alzheimer's Association (SG-14-312644) and the Fidelity Biosciences Research Initiative to MAPV; the BrightFocus Foundation (A2011048 to MAPV). NIA-LOAD Family-Based Study supported the collection of samples used in this study through NIH grants U24 AG026395 and R01 AG041797 and the MIRAGE cohort was supported through the NIH grants R01 AG025259 and R01 AG048927. We thank contributors, including the Alzheimer's disease Centers who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. Study design: HNC, BWK, JLH, MAPV; Sample collection: MLC, JMV, RMC, LAF, JLH, MAPV; Whole exome sequencing and Sanger sequencing: SR, PLW; Sequencing data analysis: HNC, BWK, KLHN, SR, MAK, JRG, ERM, GWB, MAPV; Statistical analysis: BWK, KLHN, JMJ, MAPV; Preparation of manuscript: HNC, BWK. The authors jointly discussed the experimental results throughout the duration of the study. All authors read and approved the final manuscript.

References

1. Qiu C, Kivipelto M, Von Strauss E. Epidemiology of Alzheimer's disease: Occurrence, determinants and strategies toward intervention. Dialogues Clin Neurosci. 2009; 11:111–128. [PubMed: 19585947]
2. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, et al. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry. 2006; 63:168–174. [PubMed: 16461860]
3. Van Cauwenberghe C, Van Broeckhoven C, Sleegers K. The genetic landscape of Alzheimer disease: Clinical implications and perspectives. Genet Med. 2015; 18:421–430. [PubMed: 26312828]
4. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993; 261:921–923. [PubMed: 8346443]
5. Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, et al. ApolipoproteinE: High-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A. 1993; 90:1977–1981. [PubMed: 8446617]
6. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013; 45:1452–1458. [PubMed: 24162737]
7. Pottier C, Hannequin D, Coutant S, Rovelet-Lecrux A, Wallon D, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. Mol Psychiatry. 2012; 17:875–879. [PubMed: 22472873]
8. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, et al. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013; 368:117–127. [PubMed: 23150934]
9. Guerreiro RJ, Lohmann E, Kinsella E, Bras JM, Luu N, et al. Exome sequencing reveals an unexpected genetic cause of disease: NOTCH3 mutation in a Turkish family with Alzheimer's disease. Neurobiol Aging. 2012; 33:1008.e17–1008.e23.
10. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med. 2013; 368:107–116. [PubMed: 23150908]
11. Benitez BA, Jin SC, Guerreiro R, Graham R, Lord J, et al. Missense variant in TREML2 protects against Alzheimer's disease. Neurobiol Aging. 2014; 35:1510.e19–1510.e26.
12. Carney RM, Kohli MA, Kunkle BW, Naj AC, Gilbert JR, et al. Parkinsonism and distinct dementia patterns in a family with the MAPT R406W mutation. Alzheimers Dement. 2014; 10:360–365. [PubMed: 23727082]

13. Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature. 2014; 505:550–554. [PubMed: 24336208]
14. Dolzhanskaya N, Gonzalez MA, Speriani F, Stefl S, Messing J, et al. A novel p.Leu(381)Phe mutation in presenilin 1 is associated with very early onset and unusually fast progressing dementia as well as lysosomal inclusions typically seen in Kufs disease. J Alzheimers Dis. 2014; 39:23–27. [PubMed: 24121961]
15. Logue MW, Schu M, Vardarajan BN, Farrell J, Bennett DA, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. Alzheimers Dement. 2014; 10:609–618. e11. [PubMed: 25172201]
16. Sassi C, Guerreiro R, Gibbs R, Ding J, Lupton MK, et al. Investigating the role of rare coding variability in Mendelian dementia genes (APP, PSEN1, PSEN2, GRN, MAPT and PRNP) in late-onset Alzheimer's disease. Neurobiol Aging. 2014; 35:2881.e1–2881.e6.
17. Wetzel-Smith MK, Hunkapiller J, Bhangale TR, Srinivasan K, Maloney JA, et al. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. Nat Med. 2014; 20:1452–1457. [PubMed: 25419706]
18. Kohli MA, Cukier HN, Hamilton-Nelson KL, Rolati S, Kunkle BW, et al. Segregation of a rare TTC3 variant in an extended family with late-onset Alzheimer disease. Neurol Genet. 2016; 2:e41. [PubMed: 27066578]
19. Velez JI, Lopera F, Sepulveda-FAlla D, Patel HR, Jphar AS, et al. APOE*E2 allele delays age of onset in PSEN1 E280A Alzheimer's disease. Mol Psychiatry. 2016; 21:916–924. [PubMed: 26619808]
20. Ehling R, Nosková L, Stránecký V, Hartmannová H, Přistoupilová A, et al. Cerebellar dysfunction in a family harboring the PSEN1 mutation co-segregating with a cathepsin D variant p.A58V. J Neurol Sci. 2013; 326:75–82. [PubMed: 23415546]
21. Louwersheimer E, Cohn-Hokke PE, Pijnenburg YA, Weiss MM, et al. Rare genetic variant in SORL1 may increase penetrance of Alzheimer's disease in a family with several generations of APOE-ε4 Homozygosity. J Alzheimers Dis. 2017; 56:63–74. [PubMed: 27911290]
22. Kunkle BW, Jaworski J, Barral S, Vardarajan B, Beecham GW, et al. Genome-wide linkage analyses of non-Hispanic white families identify novel loci for familial late-onset Alzheimer's disease. Alzheimers Dement. 2016; 12:2–10. [PubMed: 26365416]
23. McKhann G, Drachman D, Folstein M, Katzman R, Price D, et al. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984; 34:939–944. [PubMed: 6610841]
24. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institutes on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011; 7:263–269. [PubMed: 21514250]
25. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, et al. The diagnosis of mild cognitive impairment 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; 7:270–279. [PubMed: 21514249]
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows- Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]
27. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]
28. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–498. [PubMed: 21478889]
29. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. The human genome browser at UCSC. Genome Res. 2002; 12:996–1006. [PubMed: 12045153]

30. Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC genome browser. Bioinformatics. 2014; 30:1003–1005. [PubMed: 24227676]
31. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009; 25:2865–2871. [PubMed: 19561018]
32. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]
33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
34. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002; 30:97–101. [PubMed: 11731797]
35. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]
36. Li MX, Gui HS, Kwan JS, Bao SY, Sham PC. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. Nucleic Acids Res. 2012; 40:e53. [PubMed: 22241780]
37. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. Bioinformatics. 2015; 31:2202–2204. [PubMed: 25701572]
38. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014; 46:310–315. [PubMed: 24487276]
39. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: An accessible system for testing SNV novelty. Bioinformatics. 2011; 27:3216–3217. [PubMed: 21965822]
40. Shearer AE, Eppsteiner RW, Booth KT, Ephraim SS, Gurrola J, et al. Utilizing ethnic-specific differences in minor allele frequency to recategorize reported pathogenic deafness variants. Am J Hum Genet. 2014; 95:445–453. [PubMed: 25262649]
41. Richards S, Aziz N, Bale S, Bick D, Das S, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015; 17:405–424. [PubMed: 25741868]
42. Green RC, Cupples LA, Go R, Benke KS, Edeki T, et al. Risk of dementia among white and African American relatives of patients with Alzheimer disease. JAMA. 2002; 287:329–336. [PubMed: 11790212]
43. Lee JH, Cheng R, Graff-Radford N, Foroud T, Mayeux R, et al. Analysis of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: Implication of additional loci. Arch Neurol. 2008; 65:1518–1526. [PubMed: 19001172]
44. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]
45. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 2010; 87:139–145. [PubMed: 20598278]
46. Ryan EJ, Marshall AJ, Magaletti D, Floyd H, Draves KE, et al. Dendritic cell-associated lectin-1: A novel dendritic cell-associated, C-type lectin-like molecule enhances T cell secretion of IL-4. J Immunol. 2002; 169:5638–5648. [PubMed: 12421943]
47. Gonzalez-Dominguez E, Samaniego R, Flores-Sevilla JL, Campos-Campos SF, Gomez-Campos G, et al. CD163L1 and CLEC5A discriminate subsets of human resident and inflammatory macrophages *in vivo*. J Leukoc Biol. 2015; 98:453–466. [PubMed: 25877931]
48. Moeller JB, Nielson MJ, Reichhardt MP, Schlosser A, Sorensen GL, et al. CD163-L1 is an endocytic macrophage protein strongly regulated by mediators in the inflammatory response. J Immunol. 2012; 188:2399–2409. [PubMed: 22279103]
49. Minter MR, Taylor JM, Crack PJ. The contribution of neuroflammation to amyloid toxicity in Alzheimer'd disease. J Neurochem. 2016; 136:457–474. [PubMed: 26509334]

50. Kask M, Pruunsild P, Timmusk T. Bidirectional transcription from human LRRTM2/CTNNA1 and LRRTm1/CTNNA2 gene loci leads to expression of N-terminally truncated CTNNA1 and CTNNA2 isoforms. Biochem Biophys Res Commun. 2011; 411:56–61. [PubMed: 21708131]
51. Kolakowski LF, O'Neill GP, Howard AD, Broussard SR, Sullivan KA, et al. Molecular characterization and expression of cloned human galanin receptors GALR2 and GALR3. J Neurochem. 1998; 71:2239–2251. [PubMed: 9832121]
52. Steiner RA, Hohmann JG, Holmes A, Wrenn CC, Cadd G, et al. Galanin transgenic mice display cognitive and neurochemical deficits characteristic of Alzheimer's disease. Proc Natl Acad Sci U S A. 2001; 98:4184–4189. [PubMed: 11259657]
53. Wu F, Yao PJ. Clathrin-mediated endocytosis and Alzheimer's disease: An update. Ageing Res Rev. 2009; 8:147–149. [PubMed: 19491039]
54. Garcia-Escudero V, Martin-Maestro P, Perry G, Avila J. Deconstructing mitochondrial dysfunction in Alzheimer disease. Oxid Med Cell Longev. 2013; 2013:162152. [PubMed: 23840916]
55. Maystadt I, Rezsohazy R, Barkats M, Duque S, Vannuffel P, et al. The nuclear factor kappaB-activator gene PLEKHG5 is mutated in a form of autosomal recessive lower motor neuron disease with childhood onset. Am J Hum Genet. 2007; 81:67–76. [PubMed: 17564964]
56. Azzedine H, Zavadakova P, Plante-Bordeneuve V, Vaz Pato M, Pinto N, et al. PLEKHG5 deficiency leads to an intermediate form of autosomal-recessive Charcot-Marie-Tooth disease. Hum Mol Genet. 2013; 22:4224–4232. [PubMed: 23777631]
57. Christopherson KS, Ullian EM, Stokes CC, Mullowney CE, Hell JW, et al. Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. Cell. 2005; 120:421–433. [PubMed: 15707899]
58. Karch CM, Jeng AT, Nowotny P, Cady J, Cruchaga C, et al. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. PLoS ONE. 2012; 7:e50976. [PubMed: 23226438]
59. Vardarajan BN, Zhang Y, Lee JH, Cheng R, Bohm C, et al. Coding mutations in SORL1 and Alzheimer disease. Ann Neurol. 2015; 77:215–227. [PubMed: 25382023]
60. Benitez BA, Karch CM, Cai Y, Jin SC, Cooper B, et al. The PSEN1, p.E318G variant increases the risk of Alzheimer's disease in APOE-epsilon4 carriers. PLoS Genet. 2013; 9:e1003685. [PubMed: 23990795]
61. Liu Y, Yu JT, Wang HF, Hao XK, Yang YF, et al. Association between NME8 locus polymorphism and cognitive decline, cerebrospinal fluid and neuroimaging biomarkers in Alzheimer's disease. PLoS ONE. 2014; 9:e114777. [PubMed: 25486118]

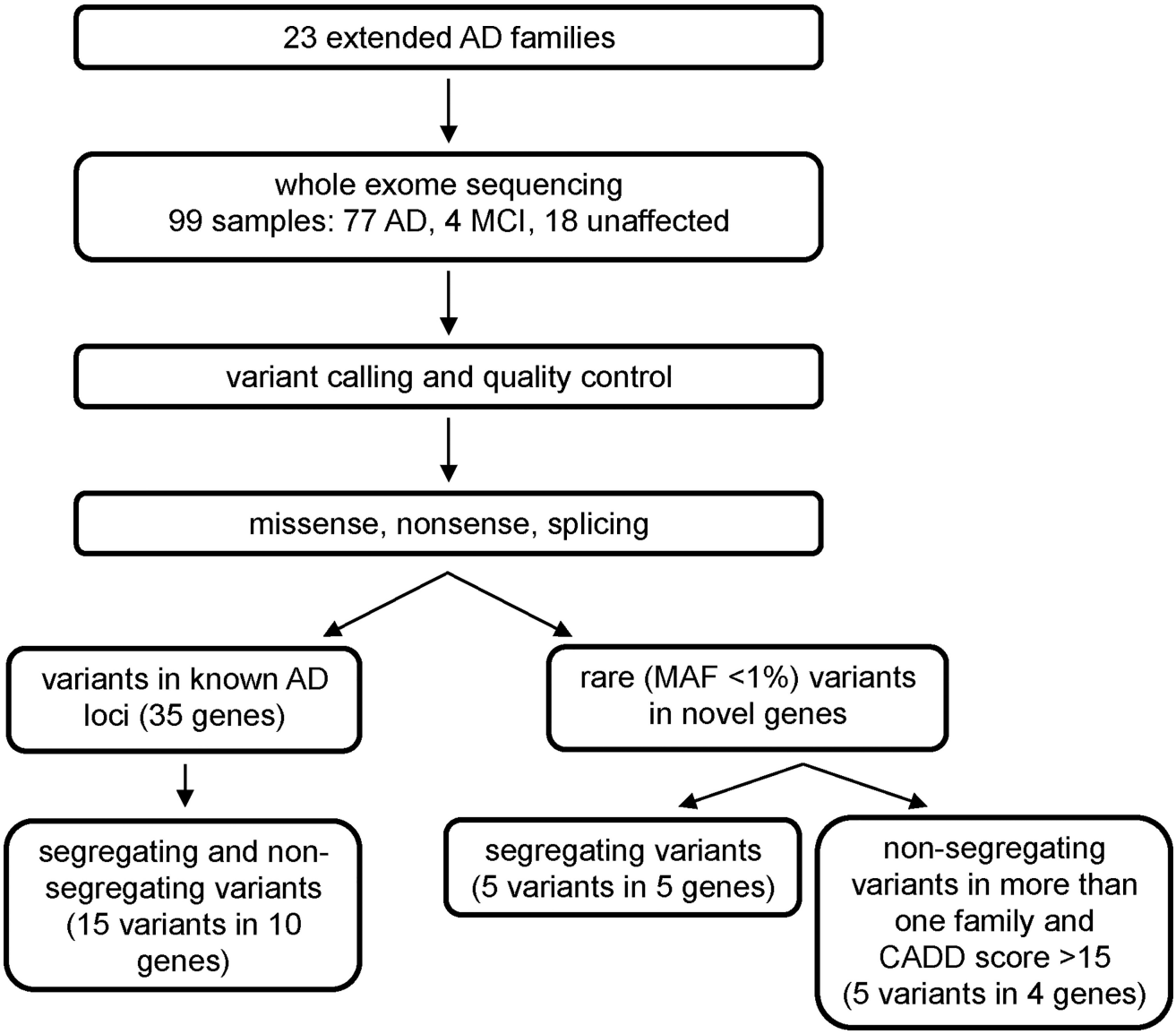


Figure 1.
Study design. Strategy for processing the samples and prioritizing the variants that were resulting from whole exome sequencing.

Table 1

Known AD genes and loci with rare, potentially damaging variants.

Gene	Chr	Position (hg38)	Nucleotide	Amino acid	dbSNP	Kaviar MAF	CADD score	Family	Affected individuals with variant/ total affected sequenced
AKAP9	7	92002147	G>A	Glu756Lys	rs202091548	0.00008	27.8	191	1/2
AKAP9	7	92017092	G>A	Arg1288Gln	rs146797353	0.00822	6.2	419	2/2
CD33	19	51225851	CCCGG>C	Gly210Thrfs*2	rs201074739	0.01339	-	2349	1/2
CR1	1	207618089	A>G	Lys2308Arg	rs41274770	0.01463	11.9	1893	2/3
EPHA1	7	143398060	C>T	Arg492Gln	rs11768549	0.01214	17.39	701	1/3
INPP5D	2	233125865	G>A	Arg157Gln	rs200834931	0.00139	17.75	1399	2/2
INPP5D	2	233206711	C>A	Ala994Asp	rs187622749	0.00433	22.8	2349	2/2
NME8	7	37884315	G>A	Arg336His	rs62001869	0.01436	6.08	1893	3/3
PSEN1	14	73206470	A>G	Glu318Gly	rs17125721	0.01423	16.92	419	1/2
SORL1	11	121543625	C>T	Thr588Ile	rs752726649	0.00001	32	191	2/2
SORL1	11	121627591	C>T	Thr2134Met	rs142884576	0.00023	28.6	1240	1/2
TREM2	6	41161469	C>T	Arg92His	rs143332484	0.00791	11.11	1893	1/3
UNC5C	4	95170263	C>T	Ala860Thr	rs34585936	0.01808	33	191	2/2
UNC5C	4	95170263	C>T	Ala860Thr	rs34585936	0.01808	33	2119	1/2
UNC5C	4	95202928	G>A	Pro666Ser	rs760453427	0.00001	20.2	191	1/2

J Alzheimers Dis Parkinsonism. Author manuscript; available in PMC 2017 November 22.

Table 2

Families with segregating, rare, potentially damaging variants in high LOD regions.

Family	Affected individuals with variant	LOD score	Gene	Chr	Position (hg38)	Nucleotide	Amino acid	dbSNP	Kaviar score	CADD score
757	9	2.95	*CD163L1*	12	7369477	T>C	Thr1317Ala	rs150384982	0.00137	3.73
757	9	2.95	*CLECL1*	12	9722727	T>C	Thr135Ala	rs118152239	0.00769	0.03
911	7	2.36	*CTNNA1*	5	138824559	G>C	Gln206His	rs150893072	0.0043	23.2
1201	5	2.22	*GALR3*	22	37823563	C>G	Pro53Ala	rs78650836	0.00328	10.31
1201	5	2.22	*MIEF1*	22	39512414	C>T	Arg169Trp	rs2232088	0.00525	34

Table 3

Genes with rare, potentially damaging variants in more than one family.

Gene	Family	Chr	Position (hg38)	Nucleotide	Amino acid	dbSNP	Kaviar score	CADD score	Affected individuals with variant/ total affecteds sequenced
DAAM2	191	6	39861001	A>G	Tyr128Cys	rs201047462	0.00017	28.00	2/2
	716	6	39884013	C>T	Arg680Trp	rs200964833	0.00019	35.00	3/3
MKL2	1893	16	14245640	T>C	Ser398Pro	rs113935526	0.00479	19.20	3/3
	26044	16	14245640	T>C	Ser398Pro	rs113935526	0.00479	19.20	2/3
PLEKHG5	803	1	6496525	G>A	Pro40Ser	rs201669114	0.00148	27.00	2/2
	1008	1	6496525	G>A	Pro40Ser	rs201669114	0.00148	27.00	2/2
THBS2	1240	6	169220312	C>A	Val1133Phe	rs112533700	0.00017	29.50	2/2
	1893	6	169220312	C>A	Val1133Phe	rs112533700	0.00017	29.50	3/3

J Alzheimers Dis Parkinsonism. Author manuscript; available in PMC 2017 November 22.

JAMA Neurology | Original Investigation

Early-Onset Alzheimer Disease and Candidate Risk Genes Involved in Endolysosomal Transport

Brian W. Kunkle, PhD, MPH; Badri N. Vardarajan, PhD; Adam C. Naj, PhD; Patrice L. Whitehead, BS; Sophie Rolati, MS; Susan Slifer, MS; Regina M. Carney, MD; Michael L. Cuccaro, PhD; Jeffery M. Vance, MD, PhD; John R. Gilbert, PhD; Li-San Wang, PhD; Lindsay A. Farrer, PhD; Christiane Reitz, MD, PhD; Jonathan L. Haines, PhD; Gary W. Beecham, PhD; Eden R. Martin, PhD; Gerard D. Schellenberg, PhD; Richard P. Mayeux, MD, MSc; Margaret A. Pericak-Vance, PhD

IMPORTANCE Mutations in *APP*, *PSEN1*, and *PSEN2* lead to early-onset Alzheimer disease (EOAD) but account for only approximately 11% of EOAD overall, leaving most of the genetic risk for the most severe form of Alzheimer disease unexplained. This extreme phenotype likely harbors highly penetrant risk variants, making it primed for discovery of novel risk genes and pathways for AD.

OBJECTIVE To search for rare variants contributing to the risk for EOAD.

DESIGN, SETTING, AND PARTICIPANTS In this case-control study, whole-exome sequencing (WES) was performed in 51 non-Hispanic white (NHW) patients with EOAD (age at onset <65 years) and 19 Caribbean Hispanic families previously screened as negative for established *APP*, *PSEN1*, and *PSEN2* causal variants. Participants were recruited from John P. Hussman Institute for Human Genomics, Case Western Reserve University, and Columbia University. Rare, deleterious, nonsynonymous, or loss-of-function variants were filtered to identify variants in known and suspected AD genes, variants in multiple unrelated NHW patients, variants present in 19 Hispanic EOAD WES families, and genes with variants in multiple unrelated NHW patients. These variants/genes were tested for association in an independent cohort of 1524 patients with EOAD, 7046 patients with late-onset AD (LOAD), and 7001 cognitively intact controls (age at examination, >65 years) from the Alzheimer's Disease Genetics Consortium. The study was conducted from January 21, 2013, to October 13, 2016.

MAIN OUTCOMES AND MEASURES Alzheimer disease diagnosed according to standard National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer Disease and Related Disorders Association criteria. Association between Alzheimer disease and genetic variants and genes was measured using logistic regression and sequence kernel association test–optimal gene tests, respectively.

RESULTS Of the 1524 NHW patients with EOAD, 765 (50.2%) were women and mean (SD) age was 60.0 (4.9) years; of the 7046 NHW patients with LOAD, 4171 (59.2%) were women and mean (SD) age was 77.4 (8.6) years; and of the 7001 NHW controls, 4215 (60.2%) were women and mean (SD) age was 77.4 (8.6) years. The gene *PSD2*, for which multiple unrelated NHW cases had rare missense variants, was significantly associated with EOAD ($P = 2.05 \times 10^{-6}$; Bonferroni-corrected *P* value [BP] $= 1.3 \times 10^{-3}$) and LOAD ($P = 6.22 \times 10^{-6}$; BP $= 4.1 \times 10^{-3}$). A missense variant in *TCIRG1*, present in a NHW patient and segregating in 3 cases of a Hispanic family, was more frequent in EOAD cases (odds ratio [OR], 2.13; 95% CI, 0.99-4.55; $P = .06$; BP = 0.413), and significantly associated with LOAD (OR, 2.23; 95% CI, 1.37-3.62; $P = 7.2 \times 10^{-4}$; BP $= 5.0 \times 10^{-3}$). A missense variant in the LOAD risk gene *RIN3* showed suggestive evidence of association with EOAD after Bonferroni correction (OR, 4.56; 95% CI, 1.26-16.48; $P = .02$, BP = 0.091). In addition, a missense variant in *RUFY1* identified in 2 NHW EOAD cases showed suggestive evidence of an association with EOAD as well (OR, 18.63; 95% CI, 1.62-213.45; $P = .003$; BP = 0.129).

CONCLUSIONS AND RELEVANCE The genes *PSD2*, *TCIRG1*, *RIN3*, and *RUFY1* all may be involved in endolysosomal transport—a process known to be important to development of AD. Furthermore, this study identified shared risk genes between EOAD and LOAD similar to previously reported genes, such as *SORL1*, *PSEN2*, and *TREM2*.

JAMA Neurol. 2017;74(9):1113-1122. doi:10.1001/jamaneurol.2017.1518
Published online July 24, 2017.

Supplemental content

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

Corresponding Author: Margaret A. Pericak-Vance, PhD, John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, 1501 NW 10th Ave, Miami, FL 33136 (mpericak@med.miami.edu).

Early-onset Alzheimer disease (EOAD), commonly defined as having age-at-onset (AAO) AD before age 65 years, accounts for approximately 10% of all cases of AD.[1] Rare mutations (minor allele frequency <0.001) in *APP* (351 Entrez Gene), *PSEN1* (5663 Entrez Gene), and *PSEN2* (5664 Entrez Gene) are the main genetic risk factors for EOAD,[2] which has a prevalence estimated as 54 per 100 000 individuals aged 30 to 65 years, and 98 per 100 000 of those aged 45 to 64 years.[3] The highly penetrant mutations in these genes account for 60% to 70% of familial EOAD and 5% to 10% of EOAD overall, leaving the majority of genetic risk for this most severe form of AD unexplained. Identifying additional loci harboring highly penetrant, rare risk variants for EOAD has been challenging, although research implicates late-onset AD (LOAD) risk genes, such as *SORL1* (6653 Entrez Gene)[4] and *TREM2* (54209 Entrez Gene),[5-7] in the development of EOAD, highlighting the potential for shared genes and pathways between the early and late forms of AD. This shared genetic architecture is likely, given their similar pathology[8,9] and the arbitrary nature of the commonly used criterion of AAO younger than 65 years delineating EOAD from LOAD.

Analysis of EOAD, which has a strong genetic component, should enhance identification of additional AD risk loci as these cases likely harbor rare, highly penetrant risk variants for disease, whereas the more common late-onset phenotype is expected to have a more complex genetic architecture.[10-12] Following this hypothesis, we performed whole-exome sequencing (WES) in 51 non-Hispanic white (NHW) individuals with EOAD (previously screened negative for known EOAD risk variants in *APP*, *PSEN1*, and *PSEN2*) to search for rare variants contributing to the risk for EOAD. Variant filtering for heterozygous functional rare variants was performed to identify high-priority variants and genes. Identified candidate variants and genes underwent additional testing in large EOAD and LOAD case-control data sets.

Key Points

Question Are there additional rare variants that contribute to the risk of early-onset Alzheimer disease?

Findings This case-control study of whole-exome sequencing of 93 patients within early-onset Alzheimer disease cases followed by testing of candidate risk variants found an association between several endolysosomal-related variants and genes with early-onset and late-onset Alzheimer disease. These included suggestive evidence of association for variants in the genes *RIN3* and *RUFY1*, a significant association with a variant in *TCIRG1*, and a significant gene-based association with *PSD2*.

Meaning This study highlights the involvement of additional endolysosomal genes in the risk of both early- and late-onset Alzheimer disease.

Methods

WES of EOAD Cases

Selection of EOAD Cases for Sequencing

Familial and sporadic NHW patients with EOAD with AAO younger than 65 years (mean, 54 years; range, 44-64 years) and thus potentially fitting the profile of either *APP*, *PSEN1*, or *PSEN2* cases were sequenced for established mutations in these genes on ascertainment to eliminate individuals with known causative genetic factors. Individuals with apolipoprotein E (*APOE*) (348 Entrez Gene) ε4/4 status, which can exhibit AAO as early as 65 years,[13] were also excluded from sequencing. In total, 51 NHW patients with EOAD were selected for WES from the John P. Hussman Institute for Human Genomics and Case Western Reserve University Alzheimer Disease Cohort (eTable 1 in the Supplement provides details). The study was conducted from January 21, 2013, to October 13, 2016. All cognitively impaired individuals, including any who changed affection status, were evaluated by the John P. Hussman Institute for Human Genomics AD clinical staff, which includes 3 of us: a geriatric psychiatrist (R.M.C.), a neurologist (J.M.V.), and a neuropsychologist (M.L.C.). In addition, 53 individuals (42 with EOAD; 11 unaffected individuals), from 19 Caribbean Hispanic families were selected for WES with mean AAO of 55 years. These families were screened for the absence of *APP*, *PSEN1*, *PSEN2*, *MAPT* (4137 Entrez Gene), and *GRN* (2896 Entrez Gene) mutations (eTable 1 in the Supplement). All affected individuals met the internationally recognized standard National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer Disease and Related Disorders Association (NINCDS/ADRA) criteria for AD.[14] The institutional review boards at University of Miami's Human Subject Research Office, Columbia University's Human Research Protection Office, and Vanderbilt University Medical Center approved all study procedures, and written informed consent was obtained from all study participants; the participants received financial compensation.

WES, Variant Calling and Quality Control

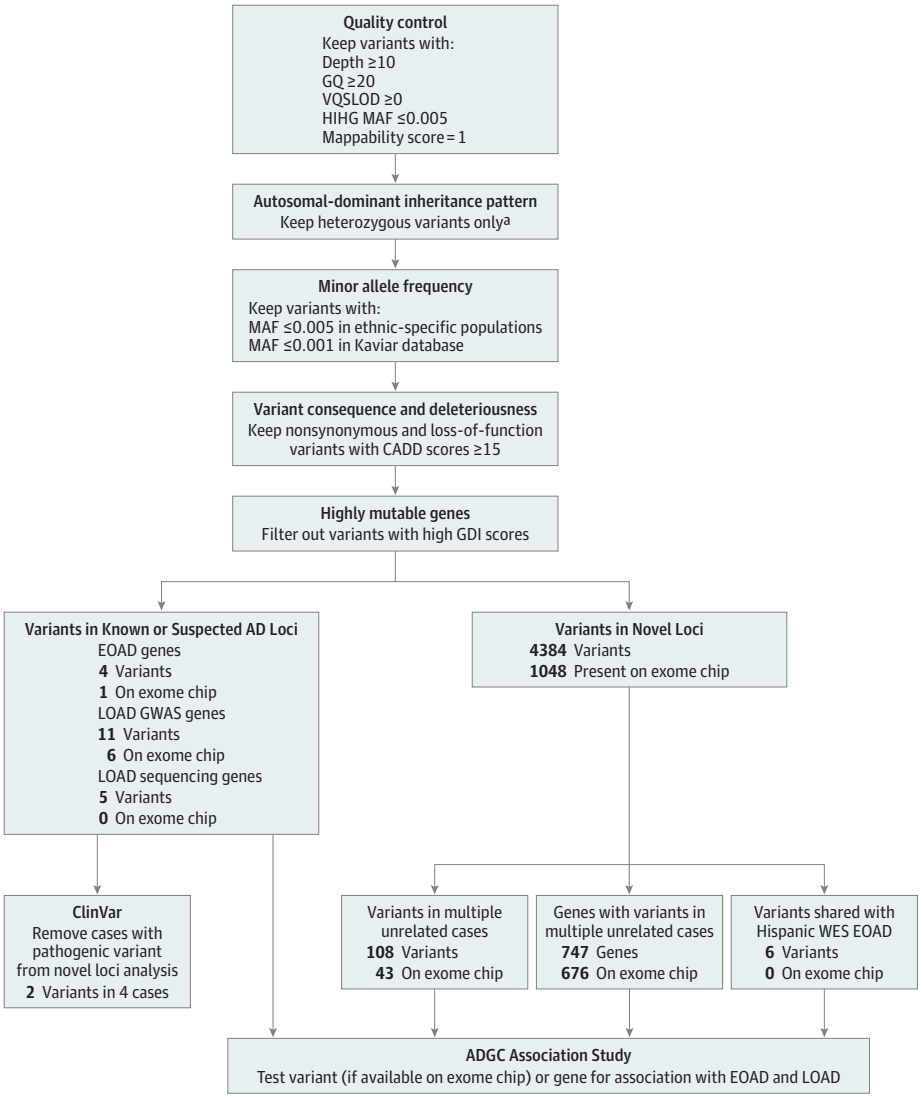
Variants were normalized using BCFTools[15,16] and variants with read depth less than 10, variant quality score log-odds less than 0, genotype quality less than 20, and 20 base pair genome mappability scores less than 1 from the Duke Uniqueness Track[17] were removed from further analysis. Reported variants were confirmed with Sanger sequencing. Further details of the WES protocol can be found in the eMethods in the Supplement.

WES Variant Prioritization

Variant Filtering for Rare Nonsynonymous or Loss-of-Function Variants

Filtering of WES variants prioritized for follow-up association testing was performed using KGGSeq[18] and custom perl and bash scripts. Nonsynonymous or loss-of-function variants with a global minor allele frequency of 0.001 or less that were in a heterozygous state and showed autosomal-dominant or X-linked dominant segregation in families, or existed in a heterozygous state in nonfamilial cases, were selected (**Figure 1**). Deleteriousness of these variants was assessed using Combined Annotation Dependent Depletion (CADD) scores.[19] A detailed description of the filtering steps implemented, including the assumptions behind our choice of

Figure 1. Analysis Strategy Summary, Including Prioritization of Candidate Variants and Their Testing, in Alzheimer Disease (AD) Case-Control Data Sets



ADGC indicates Alzheimer Disease Genetics Consortium; CADD, Combined Annotation Dependent Depletion; EOAD, early-onset AD; GDI, gene damaging index; GQ, genotype quality; GWAS, genome-wide association study; HIHG, John P. Hussman Institute for Human Genomics; IBD, identical by dissent; LOAD, late-onset AD; MAF, minor allele frequency; VQSLOD, variant quality score log-odds; WES, whole-exome sequencing.

[a] Use familiar segregation and IBD sharing if available.

minor allele frequency cutoff based on a maximum allele frequency calculation, is provided in the eMethods in the Supplement.

Prioritization of Variants in Known or Suspected AD Genes

After filtering on the criteria above, we first investigated variation in well-established and recently associated EOAD risk genes (*APP*, *PSEN1*, *PSEN2*, *SORL1*, and *TREM2*) and genes previously linked to EOAD (*MAPT* and *GRN*),[20-27] genes within significant genome-wide association study loci (defined as the 21 loci from Lambert et al,[28] which is the largest LOAD genome-wide association study to date), and genes with rare variants recently associated with LOAD through sequencing studies (*PLD3* [24646 Entrez Gene],[29] *UNC5C* [8633 Entrez Gene],[30] and *AKAP9* [10142 Entrez Gene][31]). Clinical significance of identified variants in these genes was assessed using the ClinVar Database on June 29, 2015.[32] Cases carrying known pathogenic variants were excluded from further analysis for novel genes.

Prioritization of Variants in Novel Candidate Genes

The remaining variants in individuals without a known pathogenic mutation were then prioritized for follow-up in the Alzheimer Disease Genetic Consortium (ADGC) EOAD association study, using the following criteria:

1. Rare nonsynonymous or loss-of-function variants in multiple NHW unrelated patients or families.
2. Rare, deleterious nonsynonymous or loss-of-function variants in the same gene in multiple NHW unrelated patients or families.
3. Variants shared between NHW patients with EOAD and 19 Hispanic EOAD families.

Table 1. Nonsynonymous or Loss-of-Function Variants in Known EOAD/Dementia Genes (Pathogenic in ClinVar and/or Segregating With MAF<0.005)

Gene	Cases, No.	AAO in Cases	Chr:Position:Allele Change	rsID	VEP	Protein Change	MAF[a]	CADD Score	Clinical Significance[b]
MAPT	3	52, 56, 61	17:44101427:C>T	rs63750424	Missense	R406W	1×10^{-5}	35	Pathogenic
MAPT	1	57	17:44039753:C>T	rs144611688	Missense	T17M	2×10^{-4}	23.6	ND
PSEN1	2	54, 56	14:73637653:C>T	rs63749824	Missense	A79V	6×10^{-6}	33	Pathogenic
PSEN1	1	50	14:73664774:C>G	rs63751019	Missense	R265G	6×10^{-6}	33	Untested
PSEN2	1	48	1:227075813:A>G	rs615757781	Startloss	M174V	5×10^{-4}	15.5	Probably nonpathogenic
SORL1	3	59-82	11:121414334:C>T	Novel	Missense	T588I	Novel	32	ND
SORL1	2	60, 69[c]	11:121461788:GC>G	Novel	Frameshift	Cys1431fs	Novel	35	ND
SORL1	3	55-84	11:121498300:C>T	rs142884576	Missense	T2134M	2×10^{-4}	28.6	ND

Abbreviations: AAO, age at onset; CADD, Combined Annotation Dependent Depletion[19]; Chr, chromosome; EOAD, early-onset Alzheimer disease; MAF, minor allele frequency; ND, no designation in ClinVar; VEP, variant-effect predictor variant consequence.[35]

[a] Kaviar Database MAF.

[b] According to ClinVar.[32]

[c] Individual has mild cognitive impairment.

Association Testing of Prioritized Variants and Genes

Prioritized variants and genes, both known and novel, were tested for association with EOAD and LOAD in a large case-control cohort from the ADGC.

Cohort Description

A total of 1524 NHW patients with EOAD (AAO≤65 years), 7046 NHW patients with LOAD (AAO>65 years), and 7001 NHW individuals serving as controls (mean age at examination, 77.4 years) from the ADGC had Illumina HumanExome1.0 array data available for analysis. Although EOAD is traditionally defined as having AAO younger than 65 years, we included individuals 65 years in the EOAD cohort to increase our sample size and power. Of the 1524 NHW patients with EOAD, 765 (50.2%) were women and mean (SD) age was 60.0 (4.9) years; of the 7046 NHW patients with LOAD, 4171 (59.2%) were women and mean age was 77.4 (8.6) years; and of the 7001 NHW controls, 4215 (60.2%) were women and mean age was 77.4 (8.6) years.

Samples from several ADGC cohorts were genotyped at 4 sites, including (1) the Center for Applied Genomics, The Children's Hospital of Pennsylvania, Philadelphia, Pennsylvania (2) Washington University, St Louis, Missouri (3) the Center for Genome Technology, John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, and (4) the Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, Manhasset, New York (Northshore) (eTable 2 in the Supplement). The ADGC received approval for analysis and use of data from the University of Pennsylvania Institutional Review Board. Participants' written or oral consents were obtained by their originating studies. A detailed description of ascertainment and the collection of genotype and phenotype data in the individual data sets of the ADGC is available in Naj et al[33] and Sims et al.[34] All affected individuals were adjudicated as possible or probable AD prior to analyses according to NINCDS/ADRDA criteria.[14]

Single Variant and Gene-Based Association

The exome array for the ADGC cohorts contains a total of 252 349 variants, a majority of which are functional rare single-nucleotide variants. A total of 158 165 variants were left for testing after quality control. Prioritized variants from our WES analysis present on the exome chip were assessed with both single-variant and gene-based analysis; genes containing prioritized variants not present on the chip were tested only through gene-based testing. Bonferroni-adjusted significance levels corrected for the number of tests per single variant or gene-based analysis category (eMethods in the Supplement provides further details).

Results

Variants in Known EOAD Genes

Several rare nonsynonymous or loss-of-function mutations in known or suspected EOAD genes were identified in our case series (**Table 1**). The *SORL1* missense mutations were identified in 3 families, 2 of which also have family members with LOAD. Two of these mutations, T588I (present in 4 affected individuals) and T2134M (present in 3 affected individuals) are mutations in the same individuals reported by Cuccaro et al.[36] A third mutation, a novel frameshift variant (Cys1431fs), was identified in 2 sisters, 1 who is *APOE*ε3/4 affected with AAO of 60 years and the other with mild cognitive impairment (age at examination, 69 years; *APOE*ε3/3).

A *PSEN1* missense mutation (A79V) previously reported in a LOAD family and classified as pathogenic by ClinVar was identified in 2 individuals with AAO of 54 years (*APOE*ε3/4) and 56 years (*APOE*ε3/4).[37] An additional *PSEN1* missense mutation was identified in a patient with AAO of 50 years (*APOE*ε3/3), and a *PSEN2* start-loss mutation (rs61757781) was present in an individual with AAO of 48 years (*APOE*ε3/3). *MAPT* R406W, previously reported in both frontotemporal dementia with parkinsonism-17[38,39] and AD,[24-26] was shared by 2 siblings and an unrelated participant. The individuals with the *MAPT* R406W and *PSEN1* A79V mutations were removed from further analyses owing to their ClinVar pathogenic classification. All other variants were novel or rated as probable nonpathogenic or untested in ClinVar (Table 1). Only 1 of the known EOAD gene variants was available from the ADGC exome chip

Table 2. Summary of Top Results for Each Prioritization Method

		Prioritization Category				
		Variants in Known Genes	Rare, Deleterious Variants in Multiple Cases	Genes With Rare, Deleterious Variants in Multiple Unrelated Cases		Shared Variants Between NHW and Hispanic Cases
Results						
Gene		*RIN3*	*RUFY1*	*PSD2*		*TCIRG1*
Chr: position: allele change		14:93022240:G>T	5:179036506:t>G	5:139216541:G>A	5:139216759:G>A	11:67810477:C>T
MAF[a]		5×10^{-4}	1×10^{-3}	6×10^{-4}	1×10^{-5}	7×10^{-4}
CADD score		23.6	16	28.5	27.4	13.2
EOAD SV	OR (95% CI)	4.56 (1.26-16.48)	18.63 (1.62-213.45)	...		2.13 (0.99-4.55)
	P value	0.02	3.8×10^{-3}	...		0.06
EOAD gene	OR (95% CI)	...	...	...		...
	P value	...	...	$2.0 \times 10^{-6,b}$		...
LOAD SV	OR (95% CI)	1.79 (0.65-4.87)	2.50 (0.28-21.73)	...		2.23 (1.37-3.62)
	P value	0.23	0.32	...		$7.2 \times 10^{-4,b}$
LOAD gene	OR (95% CI)	...	...	...		...
	P value	...	...	$6.2 \times 10^{-6,b}$		...
No. of variants tested (BPsig)		5 (.010)	43 (1.1×10^{-3})	676 (7.4×10^{-5})		7 (5.0×10^{-3})

Abbreviations: BPsig, Bonferroni *P* value significance level; CADD, Combined Annotation Dependent Depletion[19]; ellipses indicate that these tests were not available; EOAD, early-onset Alzheimer disease; LOAD, late-onset Alzheimer disease; MAF, minor allele frequency; NHW, non-Hispanic white; OR, odds ratio; SV, single variant; VEP, variant effect predictor variant consequence. [38]

[a] Kaviar Database MAF.

[b] Meets Bonferroni correction level.

study, a start-loss mutation in *PSEN2* (rs1757781), which showed no evidence for association in the EOAD or LOAD sample.

Twenty-six rare variants, 16 of which are deleterious according to CADD (eTable 3 in the Supplement), were present in known or suspected LOAD genes in our EOAD case series, including a frameshift variant in *HLA-DRB1* (3123 Entrez Gene) and missense variants in *ABCA7* (10347 Entrez Gene), *AKAP9* (10142 Entrez Gene), *CD2AP* (23607 Entrez Gene), *EPHA1* (2041 Entrez Gene), *MS4A4A* (51338 Entrez Gene), *RIN3* (79890 Entrez Gene), and *UNC5C*. Five of the known LOAD variants were on the exome chip, including a rare *RIN3* missense variant (rs150221413), which showed suggestive evidence of association with EOAD at a Bonferroni correction level of $P = .01$ for 5 variants tested (odds ratio [OR], 4.56 (95% CI, 1.26-16.48; $P = .02$ without adjustment for *APOE*, Bonferroni-corrected *P* value (BP) = 0.091; $P = .024$ with *APOE* adjustment) and, although not significant, was more frequent in LOAD cases than controls (minor allele frequency, 0.0008 and 0.0004 in cases vs controls, respectively; OR, 1.79; 95% CI, 0.65-4.87; $P = .23$, BP>.99) (**Table 2**; eTable 4 in the Supplement reports secondary model results).

Genomic control inflation factors (GIFs) and quantile-quantile plots show that our analyses are not inflated and are valid or conservative (ie, the EOAD single-variant tests) in terms of distribution of results (GIF<1.1[35]) (eFigures 1-4 in the Supplement). The quantile-quantile and GIFs for variants with allele counts of 10 or more show only slight inflation for the EOAD single-variant tests (GIF = 1.15), although finding may be due to our unbalanced case-control sample, as rescaling λ for 1000 cases and 1000 controls[40] produces a GIF of 0.92 (eFigures 5 and 6 in the Supplement).

Novel Candidate Variants and Genes

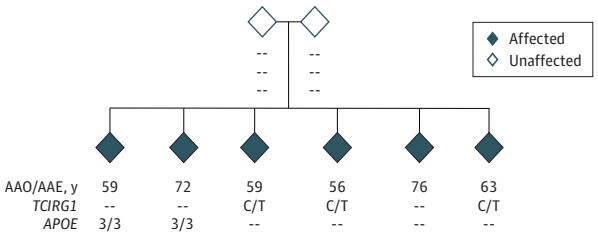
Variants Present in Multiple Unrelated Cases

After removing variants in highly mutable genes (based on high gene damage index scores), 108 rare deleterious variants in 106 genes were present in 2 or more unrelated individuals. Of these, 43 variants were testable in the ADGC exome chip data set. A missense variant in *RUFY1* (80230 Entrez Gene), present in 4 ADGC association cases and no controls, showed evidence of an association with EOAD (OR, 18.63; 95% CI, 1.62-213.45; $P = .003$, BP=.129), nearing a Bonferroni-corrected significance level of $P = 1 \times 10^{-3}$ for 43 variants tested (Table 2; eTable 4 in the Supplement provides secondary model results). All variants occurred in cases (4 EOAD and 3 LOAD), with 3 of 4 EOAD patients carrying *APOE* ε4 ($P = .28$) and 1 patient with LOAD carrying *APOE*ε4 ($P = .07$). The rarity of the variant makes it difficult to conclude whether the effect in EOAD is spurious or due to a chance correlation between *APOE*ε4 and the *RUFY1* variant. Four other variants, including a missense variant present in 2 WES EOAD cases in the gene *NAA60*, showed nominal significance in the ADGC data set (eTable 5 in the Supplement).

Genes With Variants in Multiple Unrelated Cases

Filtering to genes with rare, deleterious nonsynonymous or loss-of-function variants in multiple unrelated individuals left 747 genes, 676 of which were testable in the ADGC EOAD association study (Bonferroni-critical $P = 7.40 \times 10^{-5}$ for 676 genes tested). The gene *PSD2* (84249 Entrez Gene) met genome-wide significance in both EOAD ($P = 2.05 \times 10^{-6}$, BPsig = 1.3×10^{-3}, *APOE*-adjusted $P = 1.55 \times 10^{-5}$) and LOAD ($P = 6.22 \times 10^{-6}$, BP=4.1×10^{-3}, *APOE*-adjusted $P = 2.30 \times 10^{-4}$) cohorts when all variants in the gene were included in a gene-based test. The *APOE*-adjusted results are slightly less signifi-

Figure 2. Pedigree of Hispanic Family Segregating a Rare C>T Variant in TCIRG1



AAE indicates age at examination; AAO, age at onset; *APOE*, apolipoprotein E; and C/T, cytosine/thymine. Dashes beneath the symbols indicate that data are unavailable.

cant, likely due to smaller sample sizes of these analyses (ie, absence of *APOE* genotype for all participants) or minor correlation between *APOE* and *PSD2* risk genotypes. With restriction of the analyses to high or moderate consequence variants with CADD scores of 15 or higher, the signal for association was strengthened further (EOAD $P = 1.68 \times 10^{-6}$) (eTable 5 in the Supplement). Several additional genes (*LIN37* [55957 Entrez Gene], *SLC22A17* [51310 Entrez Gene], *LRRC16B* [90668 Entrez Gene], and *HSD17B2* [3294 Entrez Gene]) showed suggestive evidence of association with EOAD ($P < .005$) (eTable 5 in the Supplement).

Variants Present in Both NHW and Hispanic Individuals

Thirty rare missense or loss-of-function variants, 6 of which were scored as deleterious by CADD, were shared between our NHW and Hispanic WES cohorts. Seven of these variants were included on the exome chip (Bonferroni-critical $P = 7 \times 10^{-3}$ for 7 variants tested). A missense variant in *TCIRG1* (10312 Entrez Gene) (CADD Phred score, 13.2), present in NHW patients with EOAD (AAO, 57 years) and segregating in 3 Hispanic siblings with EOAD who were aged 56, 59, and 63 years (**Figure 2**), was more frequent in cases than controls (minor allele frequency, 3.2×10^{-3} and 1.4×10^{-3}, respectively) in the ADGC EOAD cohort (OR, 2.13; 95% CI, 0.99-4.55; $P = .06$, BP = .413, *APOE*-adjusted $P = .38$), and this difference was significant in the ADGC LOAD cohort (OR, 2.23; 95% CI, 1.37-3.62; $P = 7.2 \times 10^{-4}$, BP = 5.0×10^{-3}, *APOE*-adjusted $P = 2.0 \times 10^{-3}$) (Table 2; eTable 4 in the Supplement provides secondary model results). Of the16 rare nonsynonymous or loss-of-function variants prioritized in the Hispanic family, 6 are on the exome chip, with the *TCIRG1* variant being the only variant showing association with AD. Furthermore, gene-based results for genes containing the 10 other variants showed nominal association with EOAD only for the gene *COL3A1* (1281 Entrez Gene) ($P = .03$), although this gene test is not comprehensive since it relies only on the variants available on the exome chip.

Discussion

Accumulating evidence points to alterations of the endolysosomal pathway as playing key roles in AD,[41,42] with variation in several genes of the pathway recognized as risk factors for AD, including *SORL1*, *BIN1* (274 Entrez Gene), *PICALM* (8301 Entrez Gene), *RIN3*, *PTK2B* (2185 Entrez Gene), *MEF2C* (4208 Entrez Gene), and *CD2AP*.[43] Some of the earliest neuropathologic changes of AD (eg, enlargement of endosomal compartments, accumulation of phagocytic vacuoles, and lysosomal deficiencies) are endocytic in nature.[42] These abnormalities develop well before manifestation of clinical symptoms, but appear critical in the dysregulation of amyloid precursor protein processing thought to be essential in AD pathology.[44] The gene *SORL1*, which guides APP to the endocytic pathway for recycling[45] and has been linked to EOAD in several studies,[46,47] highlights a likely role for endocytosis in EOAD. Steps along the pathway include vesicle formation through membrane budding, vesicle transport, docking, cargo capture, and sorting in the early endosome, endosome maturation (late endosome), and, finally, degradation in lysosomes. It is likely that genic alterations in each of these steps contribute to the swelling of endosomal vesicles and ultimate accumulation of amyloid β in neurites that is thought to promote AD.[48-50]

Following a hypothesis that rare functional variants are responsible for EOAD, we filtered WES data on 53 NHW patients with EOAD based on consequence, deleteriousness, ethnic-specific (which have been shown to aid in the identification of true causal disease variants[51]), and population-specific minor allele frequencies. In addition to identifying several known and novel mutations in known or suspected EOAD genes (*GRN*, *MAPT*, *PSEN1*, *PSEN2*, *SORL1*, and *TREM2*), we report several candidate genes for EOAD involved in the endolysosomal pathway, including *RUFY1*, *PSD2*, *TCIRG1*, and the known LOAD risk gene *RIN3* (Table 2). These results adjusted for principal components only, but were supported for *PSD2* in secondary analyses adjusting for age, sex, and principal components ($P = 1.23 \times 10^{-3}$); however, these analyses do not show an association in the other candidate genes (eTable 4 in the Supplement), possibly due to the older mean age of the control participants. In addition, although *PSD2*, *TCIRG1*, and *RIN3* are associated with AD even with adjustment for *APOE*, the rarity of the *RUFY1* variant, which occurs only in AD, makes its evaluation in *APOE*-adjusted analysis difficult. All 4 genes participate in different steps of the endolysosomal pathway, highlighting the likelihood that alterations in many endocytic genes can increase the risk of EOAD.

The *PSD2* gene appears to play an early role through its synthesis of phospholipids critical to maturation of transport vesicles and vacuoles integral to the pathway.[52] Disturbance of the formation of these vesicles and vacuoles is critical in proper processing of endosomal debris. The importance of *PSD2* to AD in this process potentially revolves around the formation and proper maintenance of phosphatidylethanolamine, a function for which *PSD2* is essential.[53] This enzyme, which is decreased in AD brains,[54] has been shown to regulate the γ-secretase activity integral to APP processing[55] and to positively regulate autophagy and longevity in yeast.[56]

Both *RUFY1*, which binds vesicles containing the endosomal traffic regulator phosphatidylinositol-3-phosphate,[57] and *RIN3*, a known LOAD risk gene, appear to be critical to the development and regulation of the early endosome, a major site of Aβ peptide generation that is markedly enlarged within neu-

rons in AD brains.[58] While *RUFY1* binds phosphatidylinositol-3-phosphate, which is deficient in brain tissue from both humans with AD and AD mouse models,[57] it also is required for normal *RAB31* (11031 Entrez Gene) function, a *Rab5* (5868 Entrez Gene) family protein.[59] *Rab5*, a key regulator of early endosome formation, increases amyloid β production[60,61] and is stimulated and stabilized by *RIN3*.

TCIRG1, a gene that was found to share a prioritized variant between NHW and Caribbean Hispanic patients, is located in the lysosome, where it appears to be critical for acidification of vacuoles working to remove debris via the endolysosomal pathway. In AD, disturbed lysosomal degradation is of key importance in aberrant vacuole turnover.[62] Furthermore, this gene has recently been associated with absolute counts of neutrophils,[63] which are key components of innate immunity that have been linked to development of AD,[64,65] including 1 study that found 10 times more neutrophils in the brain tissue of patients with AD. The sharing of this variant in both NHW and Caribbean Hispanic populations supports the generalizability of this gene as a potential risk locus for both of these populations.

Through association analysis of our candidate genes in a large AD cohort from the ADGC, we also add to accumulating evidence pointing to overlap of risk genes involved in both EOAD and LOAD, with both *TCIRG1* and *PSD2* associated with EOAD and LOAD. This overlap of genetic architecture between the early- and late-onset forms of the disease has been previously identified for the genes *SORL1*, *TREM2*, and *PSEN1*.[66] In addition, the ε4 allele of *APOE*, the strongest genetic risk factor for LOAD, also drives the risk for EOAD in ε4/ε4 individuals with AAO of approximately 65 years and accentuates endosome pathology at early stages of AD,[58] a finding that is in line with other evidence pointing toward endolysosomal pathology occurring in the early stages of AD and promoting earlier onset of AD.[50] In addition, although some differences in neuropathology of EOAD and LOAD have been identified, many pathologic features overlap between the early and late forms of AD.[67]

Limitations

There are several limitations to our study. First, some variants prioritized in the WES analysis were not present on the exome chip, making assessment of their impact impossible in the present study. These variants should be further examined in large, case-control association studies to determine their potential risk to AD. Second, although we followed up our prioritized WES variants in a sizeable case-control sample, the power of this sample for assessment of very rare variants is limited, and replication of these results in other large case-control samples will be necessary. Finally, while several studies suggest that our top results are involved in endolysosomal transport, additional wet-laboratory studies will need to confirm that this pathway is the mechanism through which these genes increase the risk for AD.

Conclusions

Using a combined strategy of bioinformatics filtering of WES of EOAD cases, followed by testing of prioritized variants and genes in a large EOAD and LOAD cohort, we have identified several novel EOAD candidate genes, 2 of which were also associated with LOAD. Taken together, our results highlight endolysosomal alterations in multiple genes as risk factors for EOAD and point to additional genes conferring risk of both EOAD and LOAD.

ARTICLE INFORMATION

Accepted for Publication: May 19, 2017.

Published Online: July 24, 2017. doi:10.1001/jamaneurol.2017.1518

Author Affiliations: John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, Florida (Kunkle, Whitehead, Rolati, Slifer, Carney, Cuccaro, Vance, Gilbert, Beecham, Martin, Pericak-Vance); The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, New York (Vardarajan, Reitz, Mayeux); The Gertrude H. Sergievsky Center, Columbia University, New York, New York (Vardarajan, Reitz, Mayeux); Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, New York (Vardarajan, Reitz, Mayeux); Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Naj); Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Wang, Schellenberg); Department of Medicine (Biomedical Genetics), Schools of Medicine and Public Health, Boston University, Boston, Massachusetts (Farrer); Department of Neurology, Schools of Medicine and Public Health, Boston University, Boston, Massachusetts (Farrer); Department of Ophthalmology, Schools of Medicine and Public Health, Boston University, Boston, Massachusetts (Farrer); Department of Epidemiology, Schools of Medicine and Public Health, Boston University, Boston, Massachusetts (Farrer); Department of Biostatistics, Schools of Medicine and Public Health, Boston University, Boston, Massachusetts (Farrer); Institute for Computational Biology, Case Western Reserve University, Cleveland, Ohio (Haines); Department of Epidemiology, College of Physicians and Surgeons, Columbia University, New York, New York (Mayeux); Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, New York (Mayeux).

Author Contributions: Drs Vardarajan and Kunkle contributed equally to this work. Drs Kunkle and Pericak-Vance had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Kunkle, Vardarajan, Haines, Mayeux, Pericak-Vance.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Kunkle, Vardarajan, Carney, Mayeux, Pericak-Vance.
Critical revision of the manuscript for important intellectual content: Kunkle, Vardarajan, Naj, Whitehead, Rolati, Slifer, Cuccaro, Vance, Gilbert, Wang, Farrer, Reitz, Haines, Beecham, Martin, Schellenberg, Mayeux, Pericak-Vance.
Statistical analysis: Kunkle, Vardarajan, Naj, Slifer, Haines, Beecham, Martin, Pericak-Vance.
Obtained funding: Farrer, Haines, Schellenberg, Mayeux, Pericak-Vance.
Administrative, technical, or material support: Carney, Gilbert, Wang, Farrer, Schellenberg, Mayeux, Pericak-Vance.
Supervision: Carney, Gilbert, Farrer, Haines, Schellenberg, Mayeux, Pericak-Vance.

Conflict of Interest Disclosures: None reported.

Funding/Support: This study was supported by grant W81XWH-12-1-0013 from the Department of Defense. The National Institutes of Health, National Institute on Aging (NIA) supported this work through Alzheimer Disease Genetic Consortium (ADGC) grants U01 AG032984 and RC2 AG036528; samples from the National Cell Repository for Alzheimer's Disease, which receives government support under cooperative agreement grant U24 AG21886 awarded by the NIA, were used in this study. Data for this study were prepared, archived, and distributed by the National Institute on Aging Alzheimer's Disease Data Storage Site at the University of Pennsylvania (U24-AG041689-01); The Center for Genomics of Alzheimer's Disease at the University of Pennsylvania (U54-AG052427); Alzheimer's Disease Core Center at the University

of Pennsylvania (P30-AG010124); National Alzheimer's Coordinating Center, U01 AG016976; NIA LOAD (Columbia University), UF1AG047133, RF1AG015473, U01AG052410-01, U24 AG026395, U24 AG026390, R01AG041797; Banner Sun Health Research Institute P30 AG019610; Boston University, P30 AG013846, U01 AG10483, R01 CA129769, R01 MH080295, R01 AG017173, R01 AG025259, R01 AG048927, R01AG33193, R01 AG009029; Columbia University, P50 AG008702, R37 AG015473, R01 AG037212, R01 AG028786; Duke University, P30 AG028377, AG05128; Emory University, AG025688; Group Health Research Institute, U01 AG006781, U01 HG004610, U01 HG006375, U01 HG008657; Indiana University, P30 AG10133, R01 AG009956, RC2 AG036650; The Johns Hopkins University, P50 AG005146, R01 AG020688; Massachusetts General Hospital, P50 AG005134; Mayo Clinic, P50 AG016574, R01 AG032990, KL2 RR024151; Mount Sinai School of Medicine, P50 AG005138, P01 AG002219; New York University, P30 AG08051, UL1 RR029893, 5R01AG012101, 5R01AG022374, 5R01AG013616, 1RC2AG036502, 1R01AG035137; North Carolina A&T University, P20 MD000546, R01 AG28786-01A1; Northwestern University, P30 AG013854; Oregon Health & Science University, P30 AG008017, R01 AG026916; Rush University, P30 AG010161, R01 AG019085, R01 AG15819, R01 AG17917, R01 AG030146, R01 AG01101, RC2 AG036650, R01 AG22018; University of Alabama at Birmingham, P50 AG016582; University of Arizona, R01 AG031581; University of California, Davis, P30 AG010129; University of California, Irvine, P50 AG016573; University of California, Los Angeles, P50 AG016570; University of California, San Diego, P50 AG005131; University of California, San Francisco, P50 AG023501, P01 AG019724; University of Kentucky, P30 AG028383, AG05144; University of Michigan, P50 AG008671; University of Pennsylvania, P30 AG010124; University of Pittsburgh, P50 AG005133, AG030653, AG041718, AG07562, AG02365; University of Southern California, P50 AG005142; University of Texas Southwestern, P30 AG012300; University of Miami, R01 AG027944, AG010491, AG027944, AG021547, AG019757; University of Washington, P50 AG005136, R01 AG042437; University of Wisconsin, P50 AG033514; Vanderbilt University, R01 AG019085; and Washington University, P50 AG005681, P01 AG03991, P01 AG026276. The Kathleen Price Bryan Brain Bank at Duke University Medical Center is funded by NINDS grant NS39764, NIMH MH60451, and by GlaxoSmithKline. Support was also received from the Alzheimer's Association (Dr Farrer, IIRG-08-89720; Dr Pericak-Vance, IIRG-05-14147), the US Department of Veterans Affairs Administration, Office of Research and Development, Biomedical Laboratory Research Program, and BrightFocus Foundation (Dr Pericak-Vance, A2111048). The Netherlands Brain Bank (funding via numerous sources, including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, Internationale Stichting Alzheimer Onderzoek), and Institut de Neuropatologia, Servei Anatomia Patologica, Universitat de Barcelona.

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

Additional Contributions: We thank all research participants for participating in this work and their Alzheimer's Disease Genetics Consortium. Contributors included Perrie M. Adams, PhD (University of Texas Southwestern Medical Center), Erin Abner, PhD (University of Kentucky, Marilyn S. Albert, PhD (The Johns Hopkins University), Roger L. Albin, MD (University of Michigan), Liana G. Apostolova, MD (University of Indiana), Steven E. Arnold, MD (University of Pennsylvania), Sanjay Asthana, MD (University of Wisconsin), Craig S. Atwood, PhD (University of Wisconsin), Clinton T. Baldwin, PhD (Boston University), Robert C. Barber, PhD (University of North Texas Health Science Center), Michael M. Barmada, PhD (University of Pittsburgh), Lisa L. Barnes, PhD (Rush University), Sandra Barral, PhD (Columbia University), Thomas G. Beach MD, PhD (Banner Sun Health Research Institute), James T. Becker, PhD (University of Pittsburgh), Duane Beekly, BS (University of Washington), Timothy W. Behrens, MD (Genentech), David A. Bennett, MD (Rush University Medical Center), Eileen H. Bigio, MD (Northwestern University), Thomas D. Bird, MD (University of Washington), Deborah Blacker, MD (Massachusetts General Hospital/Harvard Medical School), Bradley F. Boeve, MD (Mayo Clinic), James D. Bowen, MD (Swedish Medical Center), Adam Boxer, MD, PhD (University of California, San Francisco), James R. Burke, MD, PhD (Duke University), Jeffrey M. Burns, MD, MS (University of Kansas Medical Center), Joseph D. Buxbaum, PhD (Mt Sinai School of Medicine), Nigel J. Cairns, PhD, FRCPath (Washington University), Laura B. Cantwell, MPH (University of Pennsylvania), Chuanhai Cao, PhD (University of South Florida), Chris S. Carlson, PhD (Fred Hutchinson Cancer Research Center), Cynthia M. Carlsson, MD (University of Wisconsin), Minerva M. Carrasquillo, PhD (Mayo Clinic), Helena C. Chui, MD (University of South California), Paul K. Crane, MD, PhD (University of Washington), David H. Cribbs, PhD (University of California, Irvine), Elizabeth A. Crocco, MD (University of Miami), Carlos Cruchaga, PhD (Washington University), Charles DeCarli, MD (University of California Davis), Philip L. De Jager, MD, PhD (Broad Institute), F. Yesim Demirci, MD (University of Pittsburgh), Malcolm Dick, MD (University of California Irvine), Dennis W. Dickson, MD (Mayo Clinic), Beth Dombroski, PhD (University of Pennsylvania), Rachelle S. Doody, MD, PhD (Baylor College of Medicine), Ranjan Duara, MD (Mt Sinai Medical Center), Nilufer Ertekin-Taner, MD, PhD (Mayo Clinic), Denis A. Evans, MD (Rush University Medical Center), Kelley M. Faber, MS (Indiana University), Thomas J. Fairchild, PhD (University of Texas North Health Science Center), Kenneth B. Fallon, MD (University of Alabama, Birmingham), David W. Fardo, PhD (University of Kentucky), Martin R. Farlow, MD (Indiana University), Steven Ferris, PhD (New York University), Tatiana M. Foroud, PhD (Indiana University), Matthew P. Frosch, MD, PhD (Massachusetts General Hospital), Douglas R. Galasko, MD (University of California, San Diego), Marla Gearing, PhD (Emory University), Peter St George-Hyslop, MD, FRCP (University of Toronto), Daniel H. Geschwind, MD, PhD (University of California, Los Angeles), Bernardino Ghetti, MD (Indiana University), Alison M. Goate, DPhil (Mt Sinai School of Medicine), Neill R. Graff-Radford, MD (Mayo Clinic), Robert R. Graham, PhD (Genetech), Robert C. Green, MD, MPH (Brigham and Women's Medical Center/Harvard Medical School), John H. Growdon, MD (Massachusetts General Hospital/Harvard Medical School), Hakon Hakonarson, MD, PhD (Children's Hospital of Philadelphia), Ronald L. Hamilton, MD (University of Pittsburgh), Kara L. Hamilton-Nelson, MPH (University of Miami), John Hardy, PhD (University College London), Lindy E. Harrell, MD, PhD (University of Alabama, Birmingham), Lawrence S. Honig, MD, PhD (Columbia University), Ryan M. Huebinger, PhD (University of Texas Southwestern Medical Center), Matthew J. Huentelman, PhD (Translational Genomics Research Institute), Christine M. Hulette, MD (Duke University), Bradley T. Hyman, MD, PhD (Massachusetts General Hospital/Harvard Medical School), Gail P. Jarvik, MD, PhD (University of Washington), Lee-Way Jin, MD, PhD (University of California Davis), Gyungah Jun, PhD (Boston University), M. Ilyas Kamboh, PhD (University of Pittsburgh), Mindy J. Katz, MPH (Albert Einstein College of Medicine), Anna Karydas, BA (University of California, San Francisco), John S. K. Kauwe, PhD (Brigham Young University), Jeffrey A. Kaye, MD (Oregon Health and Science University), C. Dirk Keene, MD, PhD (University of Washington), Ronald Kim, MD (University of California Irvine), Neil W. Kowall, MD (Boston University), Joel H. Kramer, PsyD (University of California, San Francisco), Walter A. Kukull, PhD (University of Washington), Frank M. LaFerla PhD (University of California Irvine), James J. Lah, MD, PhD (Emory University), Eric B. Larson, MD, PhD (University of Washington), James B. Leverenz, MD (Cleveland Clinic), Allan I. Levey, MD, PhD (Emory University), Ge Li, MD, PhD (University of Washington), Andrew P. Lieberman, MD, PhD (University of Michigan), Richard B. Lipton, MD (Albert Einstein College of Medicine), Kathryn L. Lunetta, PhD (Boston University), Constantine G. Lyketsos, MD, MHS (The Johns Hopkins University), John Malamon, MSE (University of Pennsylvania), Daniel C. Marson, JD, PhD (University of Alabama, Birmingham), Frank Martiniuk, PhD (New York University), Deborah C. Mash, PhD (University of Miami), Eliezer Masliah, MD (University of California, San Diego), Wayne C. McCormick, MD, PhD (University of Washington), Susan M. McCurry, PhD (University of Washington), Andrew N. McDavid, BA (Fred Hutchinson Cancer Research Center), Stefan McDonough, PhD (Pfizer Worldwide Research and Development), Ann C. McKee, MD (Boston University), Marsel Mesulam, MD (Northwestern University), Bruce L. Miller, MD (University of California, San Francisco), Carol A. Miller, MD (University of Southern California), Joshua W. Miller, PhD (University of California Davis), Thomas J. Montine, MD, PhD (University of Washington), John C. Morris, MD (Washington University), Shubhabrata Mukherjee, PhD (University of Washington), Amanda J. Myers, PhD (University of Miami), Sid O'Bryant, PhD (University of North Texas Health Science Center), John M. Olichney, MD (University of California Davis), Joseph E. Parisi, MD (Mayo Clinic), Amanda Partch, MS (University of Pennsylvania), Henry L. Paulson, MD, PhD (University of Michigan), William Perry, MPH (University of Miami), Elaine Peskind, MD (University of Michigan), Ronald C. Petersen, MD, PhD (Mayo Clinic), Aimee Pierce, MD (University of California Irvine), Wayne W. Poon, PhD (University of California Irvine), Huntington Potter, PhD (University of Colorado), Liming Qu, MS (University of Washington), Joseph F. Quinn, MD (Oregon

Health & Science University), Ashok Raj, MD (University of South Florida), Murray Raskind, MD (University of Washington), Eric M. Reiman, MD (Arizona Alzheimer's Consortium), Barry Reisberg, MD (New York University), Joan S. Reisch, PhD (University of Texas Southwestern Medical Center), John M. Ringman, MD (UCLA), Erik D. Roberson, MD, PhD (University of Alabama, Birmingham), Ekaterina Rogaeva, PhD (University of Toronto), Howard J. Rosen, MD (University of California, San Francisco), Roger N. Rosenberg, MD (University of Texas Southwestern), Donald R. Royall, MD (South Texas Veterans Health Administration Research Education & Clinical Center), Mark A. Sager, MD (University of Wisconsin), Mary Sano, PhD (Mt Sinai School of Medicine), Andrew J. Saykin, PsyD (Indiana University), Julie A. Schneider, MD (Rush University), Lon S. Schneider, MD (University of Southern California), William W. Seeley, MD (University of California, San Francisco), Amanda G. Smith, MD (University of South Florida), Joshua A. Sonnen, MD (University of Washington), Salvatore Spina, MD (Indiana University), Robert A. Stern, PhD (Boston University), Russell H. Swerdlow, MD (University of Kansas Medical Center), Rudolph E. Tanzi, PhD (Massachusetts General Hospital/Harvard Medical School), Tricia A. Thornton-Wells, PhD (Novartis Institutes for Biomedical Research), John Q. Trojanowski, MD, PhD (University of Pennsylvania), Juan C. Troncoso, MD (The Johns Hopkins University), Debby W. Tsuang, MD (University of Washington), Otto Valladares, MS (University of Pennsylvania), Vivianna M. Van Deerlin, MD, PhD (University of Pennsylvania), Linda J. Van Eldik, PhD (University of Kentucky), Harry V. Vinters, MD (UCLA), Jean Paul Vonsattel, MD (Columbia University), Sandra Weintraub, PhD (Northwestern University), Kathleen A. Welsh-Bohmer, PhD (Duke University), Kirk C. Wilhelmsen, MD, PhD (University of North Carolina), Jennifer Williamson, MS (Columbia University), Thomas S. Wingo, MD (Emory University), Randall L. Woltjer, MD, PhD (Oregon Health & Science University), Clinton B. Wright, MD, MS (University of Miami), Chuang-Kuo Wu, MD, PhD (Texas Tech University Health Science Center), Steven G Younkin, MD, PhD (Mayo Clinic), Chang-En Yu, PhD (University of Washington), Lei Yu, PhD (Rush University Medical Center), and Yi Zhao, MS (University of Pennsylvania). We thank D. Stephen Snyder, PhD, and Marilyn Miller, PhD (NIA), who are ex officio Alzheimer's Disease Genetics Consortium members.

REFERENCES

1. Koedam EL, Lauffer V, van der Vlies AE, van der Flier WM, Scheltens P, Pijnenburg YA. Early- versus late-onset Alzheimer's disease: more than age alone. *J Alzheimers Dis*. 2010;19(4):1401-1408.

2. Bird TD. Genetic aspects of Alzheimer disease. *Genet Med*. 2008;10(4):231-239.

3. Harvey RJ, Skelton-Robinson M, Rossor MN. The prevalence and causes of dementia in people under the age of 65 years. *J Neurol Neurosurg Psychiatry*. 2003;74(9):1206-1209.

4. Pottier C, Hannequin D, Coutant S, et al; PHRC GMAJ Collaborators. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol Psychiatry*. 2012;17(9):875-879.

5. Slattery CF, Beck JA, Harper L, et al. *R47H* TREM2 variant increases risk of typical early-onset Alzheimer's disease but not of prion or frontotemporal dementia. *Alzheimers Dement*. 2014;10(6):602-608.e4.

6. Chouery E, Delague V, Bergougnoux A, Koussa S, Serre JL, Mégarbané A. Mutations in *TREM2* lead to pure early-onset dementia without bone cysts. *Hum Mutat*. 2008;29(9):E194-E204.

7. Pottier C, Wallon D, Rousseau S, et al. *TREM2* R47H variant as a risk factor for early-onset Alzheimer's disease. *J Alzheimers Dis*. 2013;35(1):45-49.

8. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol*. 1991;82(4):239-259.

9. Jellinger KA, Bancher C. Neuropathology of Alzheimer's disease: a critical update. *J Neural Transm Suppl*. 1998;54:77-95.

10. Wingo TS, Lah JJ, Levey AI, Cutler DJ. Autosomal recessive causes likely in early-onset Alzheimer disease. *Arch Neurol*. 2012;69(1):59-64.

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

12. Pedersen NL, Posner SF, Gatz M. Multiple-threshold models for genetic influences on age of onset for Alzheimer disease: findings in Swedish twins. *Am J Med Genet*. 2001;105(8):724-728.

13. Farrer LA, Cupples LA, Haines JL, et al; APOE and Alzheimer Disease Meta Analysis Consortium. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. *JAMA*. 1997;278(16):1349-1356.

14. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34(7):939-944.

15. Li H, Handsaker B, Wysoker A, et al; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.

16. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics*. 2015;31(13):2202-2204.

17. Mappability or Uniqueness of Reference Genome of ENCODE. http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeMapability. Published 2015. Accessed September 10, 2015.

18. Li MX, Gui HS, Kwan JS, Bao SY, Sham PC. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res*. 2012;40(7):e53.

19. 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.

20. Perry DC, Lehmann M, Yokoyama JS, et al. Progranulin mutations as risk factors for Alzheimer disease. *JAMA Neurol*. 2013;70(6):774-778.

21. Lee M-J, Chen T-F, Cheng T-W, Chiu M-J. rs5848 variant of progranulin gene is a risk of Alzheimer's disease in the Taiwanese population. *Neurodegener Dis*. 2011;8(4):216-220.

22. Brouwers N, Sleegers K, Engelborghs S, et al. Genetic variability in progranulin contributes to risk for clinically diagnosed Alzheimer disease. *Neurology*. 2008;71(9):656-664.

23. Fenoglio C, Galimberti D, Cortini F, et al. Rs5848 variant influences GRN mRNA levels in brain and peripheral mononuclear cells in patients with Alzheimer's disease. *J Alzheimers Dis*. 2009;18(3):603-612.

24. Carney RM, Kohli MA, Kunkle BW, et al. Parkinsonism and distinct dementia patterns in a family with the MAPT R406W mutation. *Alzheimers Dement*. 2014;10(3):360-365.

25. Rademakers R, Dermaut B, Peeters K, et al. Tau (MAPT) mutation Arg406Trp presenting clinically with Alzheimer disease does not share a common founder in Western Europe. [1]. *Hum Mutat*. 2003;22(5):409-411.

26. Lindquist SG, Holm IE, Schwartz M, et al. Alzheimer disease-like clinical phenotype in a family with FTDP-17 caused by a MAPT R406W mutation. *Eur J Neurol*. 2008;15(4):377-385.

27. Momeni P, Pittman A, Lashley T, et al. Clinical and pathological features of an Alzheimer's disease patient with the MAPT Delta K280 mutation. *Neurobiol Aging*. 2009;30(3):388-393.

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

29. Cruchaga C, Karch CM, Jin SC, et al; UK Brain Expression Consortium; Alzheimer's Research UK Consortium. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature*. 2014;505(7484):550-554.

30. Wetzel-Smith MK, Hunkapiller J, Bhangale TR, et al; Alzheimer's Disease Genetics Consortium. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. *Nat Med*. 2014;20(12):1452-1457.

31. Logue MW, Schu M, Vardarajan BN, et al; Alzheimer Disease Genetics Consortium; Alzheimer Disease Genetics Consortium. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. *Alzheimers Dement*. 2014;10(6):609-618.e11.

32. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(Database issue):D980-D985.

33. 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.

34. Sims R, van der Lee SJ, Naj AC, et al. Rare coding variants in PLCG2, ABI3 and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease [published online July 17, 2017]. Nat Genet. http://orca.cf.ac.uk/id/eprint/98651

35. Winkler TW, Day FR, Croteau-Chonka DC, et al; Genetic Investigation of Anthropometric Traits (GIANT) Consortium. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192-1212.

36. Cuccaro ML, Carney RM, Zhang Y, et al. *SORL1* mutations and Parkinsonian features in early and late onset Alzheimer's disease. *Neurol Genet*. 2016;2(6):e116.

37. Kauwe JS, Jacquart S, Chakraverty S, et al. Extreme cerebrospinal fluid amyloid βlevels identify family with late-onset Alzheimer's disease presenilin 1 mutation. *Ann Neurol*. 2007;61(5):446-453.

38. Hutton M, Lendon CL, Rizzu P, et al. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*. 1998;393(6686):702-705.

39. Ikeuchi T, Imamura T, Kawase Y, et al. Evidence for a common founder and clinical characteristics of Japanese families with the *MAPT* R406W mutation. *Dement Geriatr Cogn Dis Extra*. 2011;1(1):267-275.

40. de Bakker PIW, Ferreira MAR, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*. 2008;17(R2):R122-R128.

41. Nixon RA. Endosome function and dysfunction in Alzheimer's disease and other neurodegenerative diseases. *Neurobiol Aging*. 2005;26(3):373-382.

42. Cataldo AM, Peterhoff CM, Troncoso JC, Gomez-Isla T, Hyman BT, Nixon RA. Endocytic pathway abnormalities precede amyloid beta deposition in sporadic Alzheimer's disease and Down syndrome: differential effects of APOE genotype and presenilin mutations. *Am J Pathol*. 2000;157(1):277-286.

43. Karch CM, Goate AM. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry*. 2015;77(1):43-51.

44. Cataldo AM, Petanceska S, Terio NB, et al. Aβ localization in abnormal endosomes: association with earliest Aβ elevations in AD and Down syndrome. *Neurobiol Aging*. 2004;25(10):1263-1272.

45. Rogaeva E, Meng Y, Lee JH, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet*. 2007;39(2):168-177.

46. Nicolas G, Charbonnier C, Wallon D, et al; CNR-MAJ collaborators. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer's disease. *Mol Psychiatry*. 2016;21(6):831-836.

47. Verheijen J, Van den Bossche T, van der Zee J, et al. A comprehensive study of the genetic impact of rare variants in SORL1 in European early-onset Alzheimer's disease. *Acta Neuropathol*. 2016;132(2):213-224.

48. Stokin GB, Lillo C, Falzone TL, et al. Axonopathy and transport deficits early in the pathogenesis of Alzheimer's disease. *Science*. 2005;307(5713):1282-1288.

49. Lee S, Sato Y, Nixon RA. Lysosomal proteolysis inhibition selectively disrupts axonal transport of degradative organelles and causes an Alzheimer's-like axonal dystrophy. *J Neurosci*. 2011;31(21):7817-7830.

50. Torres M, Jimenez S, Sanchez-Varo R, et al. Defective lysosomal proteolysis and axonal transport are early pathogenic events that worsen with age leading to increased APP metabolism and synaptic Aβ in transgenic APP/PS1 hippocampus. *Mol Neurodegener*. 2012;7:59.

51. Shearer AE, Eppsteiner RW, Booth KT, et al. Utilizing ethnic-specific differences in minor allele frequency to recategorize reported pathogenic deafness variants. *Am J Hum Genet*. 2014;95(4):445-453.

52. Bohdanowicz M, Grinstein S. Role of phospholipids in endocytosis, phagocytosis, and macropinocytosis. *Physiol Rev*. 2013;93(1):69-106.

53. Gulshan K, Shahi P, Moye-Rowley WS. Compartment-specific synthesis of phosphatidylethanolamine is required for normal heavy metal resistance. *Mol Biol Cell*. 2010;21(3):443-455.

54. Ginsberg L, Rafique S, Xuereb JH, Rapoport SI, Gershfeld NL. Disease and anatomic specificity of ethanolamine plasmalogen deficiency in Alzheimer's disease brain. *Brain Res*. 1995;698(1-2):223-226.

55. Onodera T, Futai E, Kan E, et al. Phosphatidylethanolamine plasmalogen enhances the inhibiting effect of phosphatidylethanolamine on γ-secretase activity. *J Biochem*. 2015;157(5):301-309.

56. Rockenfeller P, Koska M, Pietrocola F, et al. Phosphatidylethanolamine positively regulates autophagy and longevity. *Cell Death Differ*. 2015;22(3):499-508.

57. Morel E, Chamoun Z, Lasiecka ZM, et al. Phosphatidylinositol-3-phosphate regulates sorting and processing of amyloid precursor protein through the endosomal system. *Nat Commun*. 2013;4:2250.

58. Cataldo AM, Barnett JL, Pieroni C, Nixon RA. Increased neuronal endocytosis and protease delivery to early endosomes in sporadic Alzheimer's disease: neuropathologic evidence for a mechanism of increased β-amyloidogenesis. *J Neurosci*. 1997;17(16):6142-6151.

59. Kitagishi Y, Matsuda S. RUFY, rab and rap family proteins involved in a regulation of cell polarity and membrane trafficking. *Int J Mol Sci*. 2013;14(3):6487-6498.

60. Kajiho H, Sakurai K, Minoda T, et al. Characterization of RIN3 as a guanine nucleotide exchange factor for the Rab5 subfamily GTPase Rab31. *J Biol Chem*. 2011;286(27):24364-24373.

61. Grbovic OM, Mathews PM, Jiang Y, et al. Rab5-stimulated up-regulation of the endocytic pathway increases intracellular β-cleaved amyloid precursor protein carboxyl-terminal fragment levels and Aβ production. *J Biol Chem*. 2003;278(33):31261-31268.

62. Orr ME, Oddo S. Autophagic/lysosomal dysfunction in Alzheimer's disease. *Alzheimers Res Ther*. 2013;5(5):53-62.

63. Rosenthal EA, Makaryan V, Burt AA, et al; University of Washington, Center for Mendelian Genomics. Association between absolute neutrophil count and variation at TCIRG1: the NHLBI Exome Sequencing Project. *Genet Epidemiol*. 2016;40(6):470-474.

64. Zenaro E, Pietronigro E, Della Bianca V, et al. Neutrophils promote Alzheimer's disease-like pathology and cognitive decline via LFA-1 integrin. *Nat Med*. 2015;21(8):880-886.

65. Baik SH, Cha MY, Hyun YM, et al. Migration of neutrophils targeting amyloid plaques in Alzheimer's disease mouse model. *Neurobiol Aging*. 2014;35(6):1286-1292.

66. Cruchaga C, Haller G, Chakraverty S, et al; NIA-LOAD/NCRAD Family Study Consortium. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS One*. 2012;7(2):e31039.

67. Cacace R, Sleegers K, Van Broeckhoven C. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimers Dement*. 2016;12(6):733-748.