

ABCA7 Frameshift Deletion Associated with Alzheimer's Disease in African Americans

Holly N. Cukier, PhD^{1,2*}, Brian W. Kunkle, PhD, MPH^{1*}, Badri N. Vardarajan, PhD^{3*}, Sophie Rolati, MS¹, Kara L. Hamilton-Nelson¹, Martin A. Kohli, PhD¹, Patrice L. Whitehead, BS¹, Beth A. Dombroski, PhD⁴, Derek Van Booven, BS¹, Rosalyn Lang, PhD⁵, Derek M. Dykxhoorn, PhD^{1,6}, Lindsay A. Farrer, PhD⁷, Michael L. Cuccaro, PhD^{1,6}, Jeffery M. Vance, MD, PhD^{1,2,6}, John R. Gilbert, PhD^{1,6}, Gary W. Beecham, PhD^{1,6}, Eden R. Martin, PhD^{1,6}, Regina M. Carney, MD^{1,6}, Richard Mayeux, MD, MSc³, Gerard D. Schellenberg, PhD⁴, Goldie S. Byrd, PhD⁵, Jonathan L. Haines, PhD⁸, Margaret A. Pericak-Vance, PhD^{1,2,5♦}, for the Alzheimer's Disease Genetics Consortium.

¹John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, Florida, 33136, USA

²Department of Neurology, University of Miami, Miller School of Medicine, Miami, Florida, 33136, USA

³The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Gertrude H. Sergievsky Center, Departments of Neurology, Psychiatry, and Epidemiology, College of Physicians and Surgeons, Columbia University, New York, New York, 10032, USA

⁴Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, USA

⁵Department of Biology, North Carolina A&T State University, Greensboro, North Carolina, 27411, USA

⁶Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, Florida, 33136 USA

⁷Departments of Medicine, Neurology, Ophthalmology, Genetics & Genomics, Epidemiology, and Biostatistics, Boston University, Boston, Massachusetts, 02118, USA

⁸Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio, 44106, USA

*These authors contributed equally to the manuscript

♦Corresponding author:

Margaret A. Pericak-Vance, Ph.D.

Director, John P. Hussman Institute for Human Genomics

Executive Vice Chair, Dr. John T Macdonald Foundation Department of Human Genetics

University of Miami Miller School of Medicine

1501 NW 10th Avenue, BRB-314 (M860)

University of Miami, Miller School of Medicine Miami, Florida 33136Tel: (305) 243-2308

Fax: (305) 243-2396

Title character count: 82

Abstract word count: 246

Manuscript word count: 2,934

References: 40, e-References: 9

Figures: 3, Table: 1, e-Tables: 7

Study funding: This research was supported by grants from the National Institutes of Health [R01 AG027944, R01 AG028786 to MAPV, R01 AG019085 to JLH, P20 MD000546, R01 AG28786-01A1 to GSB, U01-AG032984, RC2-AG036528 to GDS, U24 AG026395, U24 AG026390, R01AG037212, R37 AG015473, R01AG015473 to RM, R01 AG009029 to LAF, U01-AG016976 to WAK, U24-AG021886 to TMF, R01AG009956, RC2 AG036650 to KH, UO1 AG06781, UO1 HG004610 to EL, 5R01AG20688 to MDF, P50 AG005133, AG030653 to IK, R01 AG1101, R01 AG030146, RC2 AG036650 to DAE, P30AG10161, R01AG15819, R01AG30146, R01AG17917, R01AG15819 to DAB, R01AG028786 to JJM, R01AG22018, P30AG10161 to LLB, P50AG16574, R01 032990, KL2 RR024151 to NET and NRGR, AG005138 to JDB, P50 AG05681, P01 AG03991, P01 AG026276 to AMG, and U24-AG041689 to LSW, and U19 AG047133 and UF1AG047133]; the Department of Defense [W81XWH-12-1-0013 to MAPV]; a joint grant from the Alzheimer's Association [SG-14-312644] and the Fidelity Biosciences Research Initiative to MAPV; and the BrightFocus Foundation [A2011048 to MAPV].

Search terms: Alzheimer's disease [26], all genetics [91], association studies in genetics [95]

Author contributions: HNC, BWK, BNV, MK, JG, RM, and MAPV conceived and designed the experiments. RL, MLC, JMV, RMC, GSB, JH, and MAPV acquired and assessed the HIHG samples. LAF, RM, and GDS acquired and assessed the ADGC samples. SR, PLW, and BAD performed custom capture sequencing, exome sequencing, Sanger sequencing, and TaqMan genotyping. HNC, BWK, BNV, SR, KLHN, BAD, DVB, MAK, GWB, JG, and MAPV analyzed the sequencing and TaqMan data. BWK, KLHN, and MAPV performed the statistical analysis. HNC, SR, and DMD performed RT-PCR. HNC drafted the manuscript. BWK, BNV, DMD, JMV, GWB, MAK, RM, GDS, JLH, and MAPV edited the manuscript. The authors jointly discussed the experimental results throughout the duration of the study. All authors read and approved the final manuscript.

Author disclosures:

Dr. Cukier reports no disclosures.
 Dr. Kunkle reports no disclosures.
 Dr. Vardarajan reports no disclosures.
 Ms. Rolati reports no disclosures.
 Ms. Hamilton-Nelson reports no disclosures.
 Dr. Kohli reports no disclosures.
 Ms. Whitehead reports no disclosures.
 Dr. Dombroski reports no disclosures.
 Mr. Van Booven reports no disclosures.
 Dr. Lang reports no disclosures.
 Dr. Dykxhoorn reports no disclosures.
 Dr. Farrer reports no disclosures.
 Dr. Cuccaro reports no disclosures.
 Dr. Vance reports no disclosures.
 Dr. Gilbert reports no disclosures.
 Dr. Beecham reports no disclosures.
 Dr. Martin reports no disclosures.
 Dr. Carney reports no disclosures.

Dr. Mayeux reports no disclosures.
Dr. Schellenberg reports no disclosures.
Dr. Byrd reports no disclosures.
Dr. Haines reports no disclosures.
Dr. Pericak-Vance reports no disclosures.

Abstract

Objective: To identify a causative variant(s) that may contribute to Alzheimer's disease (AD) in African Americans (AA) in the *ATP-binding cassette, sub-family 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 datasets (discovery n=1,068; replication n=1,749) and whole exome sequencing Caribbean Hispanic (CH) AD families.

Results: A 44 base pair deletion (rs142076058) was identified in all 77 risk genotype carriers, showing that the deletion is in high linkage disequilibrium with the risk allele. The deletion was assessed in a large dataset (531 cases, 527 controls) and, following adjustments for age, sex, and *APOE* status, was significantly associated with disease (p=0.0002, OR=2.13 [95% CI:1.42-3.20]). An independent dataset replicated the association (447 cases, 880 controls, p=0.0117, OR=1.65 [95% CI:1.12-2.44]), and joint analysis increased the significance (p=1.414x10⁻⁵, 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 NHW cohort. Whole exome sequencing of multiplex, Caribbean Hispanic families identified the deletion co-segregating 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.

Conclusion: This common *ABCA7* deletion could represent an ethnic specific, pathogenic alteration in Alzheimer's disease.

Introduction

Alzheimer's 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 to non-Hispanic white (NHW) populations^{1,2}.

Apolipoprotein E (APOE) was the first gene associated with AD and the $\epsilon 4$ allele confers an increased risk across populations^{3,4}. While *APOE* $\epsilon 4$ occurs more frequently in AA than NHW, paradoxically, it has a lower effect size in AA⁵⁻⁷. 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 ($p=2.21 \times 10^{-9}$, OR=1.79 [95% 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* $\epsilon 4$ 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 allele¹²⁻¹⁵. Therefore, targeted sequencing of *ABCA7* was performed to identify potential causative variants. A frameshift deletion was found associated with AD in AA, but virtually absent in NHW. Thus, this deletion potentially represents a common, ethnic-specific, and likely pathogenic alteration that confers risk to Alzheimer's disease.

Methods

Standard Protocol Approvals, Registrations, and Patient Consents

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

Sample Collection

African Americans

Individuals were ascertained for this study following 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 dataset. Each of the participants were ascertained using the protocol approved by the proper Institutional Review Boards. 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 of onset 74.0 years [SD 8.5]) as well as 529 controls (403 women and 126 men, mean age at exam 73.1 years [SD 5.4]). The complete HIHG case-control AA cohort (n=1,068) included 47 relatives, giving 1021 independent (unrelated) individuals available for analysis.

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

For both HIHG and ADGC datasets, participants underwent rigorous phenotyping and diagnostic criteria following that of the National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer’s Disease and Related Disorders Association, as previously described ^{11,15,16}. Furthermore, the cognitive status of controls was measured with either the Mini-Mental State Exam (MMSE) ¹⁷ or the Modified Mini-Mental State (3MS) ¹⁸, as well as the Clinical Dementia Rating Scale (CDR) which assesses functional decline ¹⁹. All of 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 families initially recruited as a part of the Genetic Epidemiology of Alzheimer's Disease In Hispanics (EFIGA) family study at Columbia University were utilized. A total of 49 cases and 8 unaffected relatives were involved in this study. Information about patient recruitment, demographics and clinical phenotyping have been previously published 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). 14,636 probes were chosen at a 3x density with the moderately stringent masking setting to cover 84.8% of the region. For whole exome sequencing, samples from the 19 Caribbean Hispanic families were utilized (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 50Mb 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 2x100 sequencing was performed. The sequencing data were processed using the Illumina Real Time Analysis (RTA) base calling pipeline version 1.8. The Burrows-Wheeler Aligner (BWA) 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 in order to perform Sanger sequencing for validation (Deletion-F: AAATCTTCCCGCCTTGAGAT, Deletion-R: GGAGCTTAGGGTGCAGCTC). Polymerase chain reaction (PCR) experiments were set up with 1.5 mM MgCl₂, 1.6 M betaine, and touchdown PCR 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: GCCAATATGGCAAACCATC, rs115550680-R: TCCAAAACCCTGTGATAGCC) to generate a 245 bp amplicon. PCR

reactions were set up with 2 mM MgCl₂ and touchdown PCR performed. Sequencing reactions were performed utilizing the Big Dye Terminator v3.1 (Life Technologies, Carlsbad, CA), reactions were run on a 3730xl DNA Analyzer (Life Technologies), and results 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 TaqMan SNP Genotyping Assay (Life Technologies). The *ABCA7* deletion was evaluated by a custom designed TaqMan SNP Genotyping Assay 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 analyzed with the SDS v2.4 software.

RNA Isolation and RT-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 required to have a 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: TGTTCTGCGTGTGCTGA, 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, while the allele with the deletion produces a 272 bp amplicon.

Statistical Analysis

The GENMOD program, as part of the SAS/STAT software, was utilized 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. Fisher's 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* as well as eight flanking genes and a small nuclear RNA. Samples were sequenced to an average depth of over 1000x and evaluated for single nucleotide variants (SNVs) and insertions and deletions. 1,120 SNVs were detected by sequencing with 11 variants showing different frequencies in cases and controls ($p < 0.1$, Table e-1). 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, suggesting 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, designated as HHG. Following 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 16.2% of AD cases. A subset of eight individuals were pathologically-confirmed cases of AD; two were found to carry the deletion while the remaining six did not. The AA risk allele was also genotyped in this dataset. 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 ($D'=1.000$, $r^2=0.995$, Table e-2 and e-3). The top SNPs in *ABCA7* previously reported in NHW studies were also found to be in LD 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, indicating that the genomic fragment carrying the deletion in these individuals may be of African descent.

An independent AA dataset from the Alzheimer's Disease Genetics Consortium (ADGC) was evaluated in the same manner and the deletion was again significantly 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 two cohorts resulted in an overall significant 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 dataset 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 Caribbean Hispanic families from the Dominican Republic with multiple affected AD subjects²⁷. In addition to a relatively high level of African ancestry, Caribbean Hispanics are highly inbred and have a high incidence of AD, and are thus enriched for AD genetic risk factors²⁸. We independently identified the same 44 bp deletion from whole exome sequencing of three affected individuals from a large Caribbean Hispanic 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 Alzheimer's disease (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 1MB 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. Since 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 if there was an effect of the deletion on age of onset in AD. We did not find a significant difference in the age of onset in cases with the deletion (75.6 years [SD 9.6]) compared to 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 previously reported in the Exome Aggregation Consortium (ExAC: <http://exac.broadinstitute.org>) [6, 2015], a repository of 60,706 unrelated individuals from six distinct ethnic groups. In the ExAC dataset, 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 significantly associated with AD in individuals of African ancestry. While the deletion did occur in unaffected individuals, it was found at a higher frequency in individuals with AD (15.2% of cases versus 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 datasets 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 of the deletion with the top three previously reported SNPs found a high D' across all locations, but only a significant 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 to the largest isoform that generates a protein of 2146 amino acids (Figure 1B). Since 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, two 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 phosphatidylserine²⁹. Alternatively, the shortened transcript may be subjected to nonsense-mediated decay (NMD), as was seen in the Glu709fs alteration identified in NHW³⁰. While 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 signal³⁰⁻³². 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 encodes 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 pathogenesis³⁷⁻³⁹. The *ABCA7* protein is involved in the processing of amyloid precursor protein (APP⁴⁰). In addition, evidence has shown that *ABCA7* acts in the phagocytic pathway via extracellular signal-regulated kinase signaling^{e-1,e-2}. *ABCA7* is not the only ABC transporter gene linked to AD; *ABCA1*, *ABCB1*, *ABCC1*, *ABCG1*, *ABCG2*, and *ABCG4* are all implicated in A β regulation^{e-3-e-8}. Furthermore, a study identified rare loss-of-function alterations in NHW patients

diagnosed with Parkinson's disease, including specific variants previously reported in AD individuals, demonstrating that this gene may contribute to the risk of multiple neurodegenerative disorders ^{e-9}.

Therefore, the results of this study demonstrate that there is a 44 base pair deletion in *ABCA7* that is significantly 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 Caribbean Hispanic 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 Alzheimer's disease in populations with African ancestry.

Figure Legends

Figure 1. Location of the Deletion in the *ABCA7* gene and protein. A. The *ABCA7* gene (chr19:1,040,103-1,065,571, hg38), the 44 base pair deletion (blue), the AA risk allele (blue and underlined¹¹), and three NHW risk alleles (rs3764650⁸, rs3752246⁹, rs4147929¹⁰). **B.** The wild type *ABCA7* protein (2146 amino acids) and the location of frameshift deletion (blue) identified in this study. Below, the protein the predicted to be generated from deletion would contain only two of the eleven transmembrane domains (yellow) and neither of the two AAA domains (green), but incorporate 168 aberrant amino acids (black). The remaining frameshift, nonsense, and splicing variants designated are rare alterations (<1 % MAF) previously reported in NHW populations to be associated with AD³³⁻³⁵.

Figure 2. Pedigree of an AD family from the Dominican Republic with the *ABCA7* Deletion. Family 360 has six individuals diagnosed with Alzheimer's disease, as well as two individuals presenting with mild dementia. The numbers beneath each individual in the pedigree represent the individual's sample number, the age of onset of Alzheimer's disease (for AD cases) or the age at exam, and the *APOE* genotype. Seven siblings all carry the *ABCA7* deletion as well as the AA risk allele, while individual 3 did not have either the deletion or the AA risk allele.

Figure 3. Deletion Allele Produces an RNA Transcript. A. RT-PCR from cDNA of three samples without the *ABCA7* deletion (+/+) and three 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 AA control lacking the deletion and 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.

Table 1. Association Testing of the Deletion in AA Cohorts

HIHG^a	samples with deletion/ total samples (%)	odds ratio	95% CI	Pr > Z
cases	86/531 (16.2)	2.13	1.42-3.20	0.0002
controls	49/527 (9.3)			
ADGC^b				
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	1.414x10 ⁻⁵
controls	137/1407 (9.7)			

^aHIHG – John P. Hussman Institute for Human Genomics, University of Miami

^bADGC – Alzheimer's Disease Genetics Consortium

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 ethnracial 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 USA* 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.

1922. 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, Paarman 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.

Acknowledgements

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