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) ≤ 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