African American exome sequencing identifies potential risk variants at Alzheimer disease loci

OPEN

Aurelie N'Songo, PhD

Minerva M. Carrasquillo, PhD

Xue Wang, PhD

Jeremy D. Burgess

Thuy Nguyen

Yan W. Asmann, PhD

Daniel J. Serie

Steven G. Younkin, MD, PhD

Mariet Allen, PhD

Otto Pedraza, PhD

Ranjan Duara, MD

Maria T. Greig Custo, MD

Neill R. Graff-Radford, MD

Nilüfer Ertekin-Taner, MD, PhD

Correspondence to
Dr. Ertekin-Taner:
taner.nilufer@mayo.edu

ABSTRACT

Objective: In African Americans, we sought to systematically identify coding Alzheimer disease (AD) risk variants at the previously reported AD genome-wide association study (GWAS) loci genes.

Methods: We identified coding variants within genes at the 20 published AD GWAS loci by whole-exome sequencing of 238 African American participants, validated these in 300 additional participants, and tested their association with AD risk in the combined cohort of 538 and with memory endophenotypes in 319 participants.

Results: Two *ABCA7* missense variants (rs3764647 and rs3752239) demonstrated significant association with AD risk. Variants in *MS4A6A*, *PTK2B*, and *ZCWPW1* showed significant gene-based association. In addition, coding variants in *ZCWPW1* (rs6465770) and *NME8* (rs10250905 and rs62001869) showed association with memory endophenotypes.

Conclusions: Our findings support a role for *ABCA7* missense variants in conferring AD risk in African Americans, highlight allelic heterogeneity at this locus, suggest the presence of AD-risk variants in *MS4A6A*, *PTK2B*, and *ZCWPW1*, nominate additional variants that may modulate cognition, and importantly provide a thorough screen of coding variants at AD GWAS loci that can guide future studies in this population. *Neurol Genet* 2017;3:e141; doi: 10.1212/NXG.0000000000000141

GLOSSARY

AD = Alzheimer disease; **DR** = delayed recall; **GWAS** = genome-wide association study; **IGAP** = The International Genomics of Alzheimer's Project; **IR** = immediate recall; **LD** = linkage disequilibrium; **LM** = Logical Memory; **LOAD** = late-onset Alzheimer disease; **MAF** = minor allele frequency; **OR** = odds ratio; **PR** = percent retention; **QC** = quality control; **SNP** = single nucleotide polymorphism; **VR** = Visual Reproduction; **WES** = whole-exome sequencing.

African American individuals are approximately twice as likely to develop Alzheimer disease (AD) compared to Caucasians[1–3]; however, genetic studies in African Americans are scarce.[4–6] Existing genetic studies provide evidence for genetic heterogeneity between African American and Caucasian participants[7–9] (supplemental text at Neurology.org/ng). In the largest African American genome-wide association study (GWAS) conducted in 5,896 participants, besides *APOE*, only *ABCA7* locus achieved genome-wide significance, although the most significant variant was different than that identified in Caucasians.[8] A meta-analysis of 74,046 individuals of European ancestry identified or confirmed 20 late-onset Alzheimer disease (LOAD) susceptibility loci,[10] none of which have been systematically investigated in African Americans. We performed whole-exome sequencing (WES) in an African American cohort to identify the full spectrum of coding variants in the candidate genes residing at these loci. Our study is a comprehensive coding variant discovery of LOAD candidate risk genes in African American individuals.

METHODS **Participants.** Five hundred fifty unrelated self-reported African American individuals (340 controls and 198 AD cases) were recruited for this study, of which 538 were retained for analysis after quality control (QC). All participants were recruited at Mayo Clinic in Jacksonville, FL, except 1 cognitively normal and 3 AD participants recruited at Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai, Miami, FL. WES was performed on 250 DNA samples from this cohort (figure 1). Variants of interest were validated by genotyping or Sanger sequencing (table 1).

Standard protocol approvals, registrations, and patient consents. This study was approved by the institutional review boards of all participating institutions, and all participants provided written informed consent for the conducted research.

Whole-exome sequencing. Exome capture was performed using the Agilent Sure Select V4 + UTR exome capture kit.[11] Sequencing was performed on the Illumina HiSeq 2000 platform. The study design, including sample and variant QC, is described in figure 1. Further details of these and population substructure analyses are described in the e-Methods.

Validation. Variants of interest that achieved nominal significance with AD risk in the WES cohort, either by logistic regression analysis or by Fisher exact test, were genotyped using TaqMan SNP Genotyping Assays or Sanger sequencing. Validation genotyping or sequencing was performed in the combined cohort of 340 African American controls and 198 AD cases, which were then tested for AD-risk association with the coding variants of interest. The *ABCA7* intronic variant rs3764650, which shows association with AD risk in a Caucasian LOAD GWAS,[12] was also genotyped in our entire African American cohort using a TaqMan assay.

Memory endophenotypes. Verbal and visual episodic memory associations were tested on a subset of 319 individuals (75 ADs and 244 controls) using the Logical Memory (LM) and the Visual Reproduction (VR) subtests of the Wechsler Memory Scale following standardized instructions.[13] The subset of participants included in the memory endophenotype associations were selected merely on the basis of availability of cognitive test results. For each subtest, test performance was assessed using the immediate recall (IR) raw score (LMIR and VRIR), the 30-minute delayed recall (DR) raw score (LMDR and VRDR), and the percent retention (PR) score (LMPR and

Figure 1   Study design
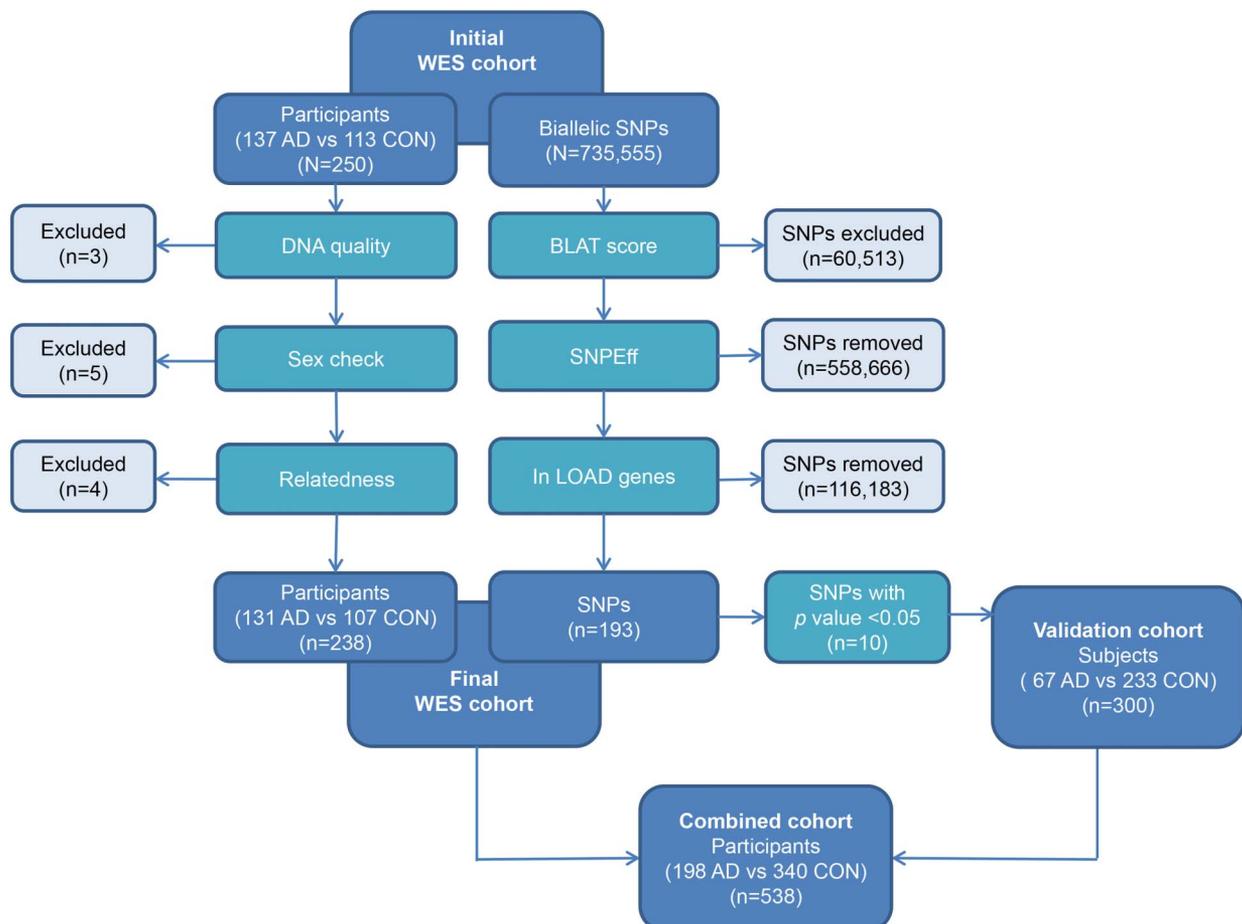
Whole-exome sequencing (WES) sample quality control (QC) led to exclusion of samples with poor sequence quality, sex ambiguity, or cryptic relatedness. WES single nucleotide polymorphism (SNP) QC led to exclusion of SNPs with a BLAST-like alignment tool (BLAT) score >1, those not resulting in an alteration of the coding sequence, and those outside the 20 late-onset Alzheimer disease (LOAD) genome-wide association study loci candidate genes. Of the remaining 193 variants tested in the final post-QC WES cohort (n = 238), 10 demonstrated association with Alzheimer disease (AD) risk with an uncorrected $p < 0.05$. After excluding 2 SNPs with strong linkage disequilibrium, the remaining 8 SNPs were confirmed by sequencing or genotyping in an independent validation cohort (n = 300). These 8 SNPs were also evaluated in the combined cohort (n = 538).

Table 1	Study cohort demographics							
	WES cohort		**Validation cohort**		**Combined cohort**		**Memory endophenotype cohort**	
	AD	Controls	AD	Controls	AD	Controls	AD	Controls
N	131	107	67	233	198	340	75	244
% Female	68.7	78.5	74.6	76.4	70.7	77.1	74.7	75.8
Mean age, y (range)	76.2 (51–99)	80.9 (51–96)	80.2 (61–96)	75.5 (33–98)	77.48 (51–99)	77.2 (33–98)	78.7 (63–96)	78.3 (60–96)
APOE ε4 frequency, %	40.8	17.7	38.8	19.3	40.1	18.8	36.7	20.7
Education, y							12.4 (7–18)	12.5 (3–20)

Abbreviations: AD = Alzheimer disease; WES = whole-exome sequencing.
The WES cohort consists of participants retained in the analyses following quality control. Validation cohort is independent of the WES cohort. Combined cohort includes participants from the WES and validation cohorts. Memory endophenotype cohort is a subset of participants from the combined cohort for whom cognitive data were available.

VRPR). Tests were administered and evaluated by Mayo Clinic neuropsychologists.

Statistical analysis. The association of the selected variants with AD risk was evaluated by multivariate logistic regression analysis adjusted for age, sex, and APOE ε4 dosage; however, variants with a total minor allele count <10 were analyzed using the Fisher exact test,[14] as implemented in PLINK.[15] We also performed multivariate logistic regression analysis adjusted for all covariates in the primary analysis plus the first 4 principal components. Haploview was used for linkage disequilibrium (LD) analysis.[16] Power analyses were performed in StatsDirect v.2.7.8. Gene-based analysis was conducted using the Sequence Kernel Optimal Unified Association Test[17] (e-Methods).

Tests of association between selected variants and the 6 memory endophenotypes (LMIR, LMDR, LMPR, VRIR, VRDR, and VRPR) were performed in R using a multivariate linear regression analysis adjusted for age, sex, APOE ε4 dosage, years of education, and diagnosis.[18] Neuropsychological test scores were obtained from the most recent neuropsychological evaluation. Age and diagnosis were defined at the time of the last neuropsychological examination.

RESULTS WES was performed on 250 African American individuals in our cohort (figure 1). Twelve participants were excluded based on QC; 3 based on poor sequence quality, 5 due to sex ambiguities, and 4 based on identity by descent (IBD) analysis. No population outliers were observed among the remaining participants based on multidimensional scaling analysis[15] (figure e-1). The final WES cohort of 131 AD and 107 controls (table 1) has 80% power to detect a variant with control minor allele frequency (MAF) of 0.10 and odds ratio (OR) = 2.96; and with MAF = 0.03 and OR = 5.35 at α = 0.05. The entire cohort composed of 238 post-QC WES participants and an additional 67 AD cases and 233 controls (198 AD and 340 controls) has 80% power to detect a variant with MAF = 0.10 and OR = 2.13; or one with MAF = 0.03 and OR = 3.3 at α = 0.05.

Following QC, 116,376 single nucleotide polymorphisms (SNPs) were retained for analysis. Of these, 193 were in LOAD GWAS candidate genes

(table e-1). AD-risk association by multivariate logistic regression analysis identified 7 variants with nominal significance (uncorrected $p < 0.05$). The Fisher exact test identified 3 additional rare SNPs (table e-2). Secondary analysis of the common variants by multivariate logistic regression analysis that also adjusted for the first 4 principal components did not substantially influence the results (tables e-1 and e-2). Five of these SNPs were in ABCA7, 2 within NME8, and 1 each in CASS4, MS4A6A, and ZCWPW1. We also tested these variants for association with AD risk by comparing their frequencies in the AD participants of our WES cohort to that in the African American population controls from the Exome Sequencing Project.[19] Four ABCA7 variants, rs3752232, rs3764647, rs3752239, and rs4147918, showed nominally significant association in this analysis with the same direction of effect as that observed in the WES analysis.

Among the 5 ABCA7 variants, rs3764647 and rs3752232 were in tight LD ($r^2 = 0.91$); as were rs3752239 and rs4147918 ($r^2 = 0.85$) (figure 2). For each pair, the variant with the greater predicted functional impact (rs3752239) or larger OR estimate (rs3764647) was retained for further analysis. Hence, 8 SNPs were validated and subsequently tested for association with AD risk in the combined cohort. Of these, 2 missense variants in ABCA7, rs3764647 and rs3752239 continued to have nominally significant association with AD risk in addition to rs35031530 in CASS4 (table 2), although the latter departed from Hardy-Weinberg equilibrium ($p = 0.03$) in controls.

To determine whether the AD-risk association observed with the 2 ABCA7 coding variants accounted for the effect of previously reported ABCA7 GWAS variants, we first evaluated the Caucasian LOAD GWAS ABCA7 risk variant rs3764650[12] in our entire cohort. This variant was not associated with AD risk in our African American participants (multivariate logistic regression $p = 0.13$, OR = 1.28).

Figure 2   Linkage disequilibrium pattern of single nucleotide polymorphisms in *ABCA7* in the whole-exome sequencing cohort
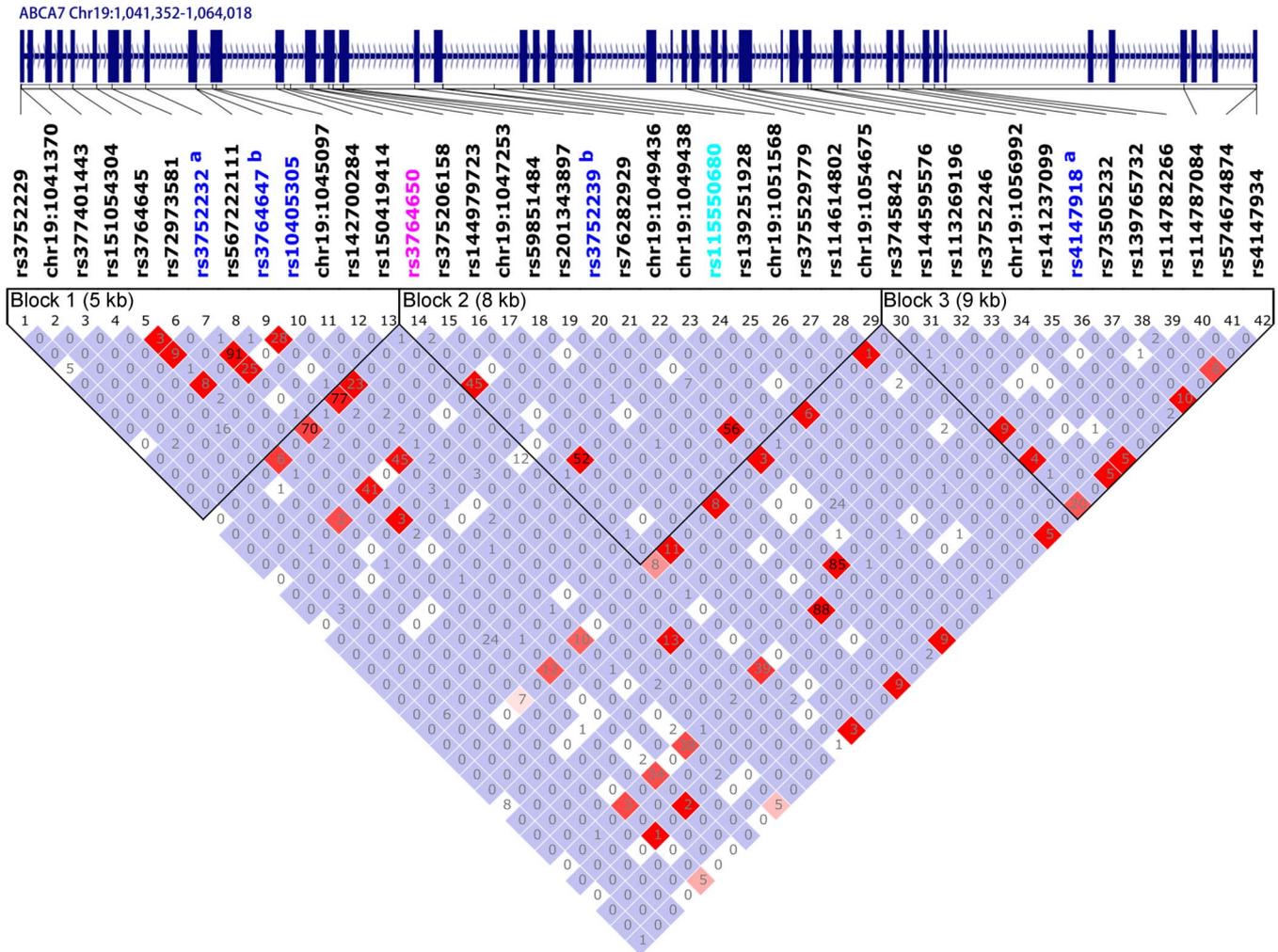
r^2 Values are shown on the plot. The Caucasian genome-wide association study (GWAS) single nucleotide polymorphism (SNP) is shown in pink font. The African American GWAS SNP was imputed in our data set ($r^2 = 0.28$) and is shown in light blue font. Given the low imputation quality, the r^2 value for this variant may not be accurate, but it is retained in this plot to show its genic location. Variants showing nominal association with Alzheimer disease risk in our data set are represented in navy blue font. [a]These 2 *ABCA7* variants were in strong linkage disequilibrium (LD) with 2 other *ABCA7* variants ([b]) identified in the whole-exome sequencing and were therefore not included in the validation. [b]These 2 *ABCA7* variants that were in strong LD with 2 others ([a]) were retained either because they had a larger odds ratio estimate or greater functional potential.

However, rs3764650 is in LD with the *ABCA7* coding variant rs3764647 ($r^2 = 0.77$) (figure 2). Conditional analysis adjusting for rs3764650 genotypes abolished the AD-risk association of rs3764647 (conditional $p = 0.198$, conditional OR = 1.54). In contrast, there is no LD between rs3764650 and the other significant *ABCA7* coding variant rs3752239, which remains nominally significant even after accounting for rs3764650 ($p = 0.047$, OR = 4.65).

Previous AD GWAS in African American participants identified the *ABCA7* intronic variant rs115550680 to be significantly associated with AD risk.[8] This intronic variant is located at position 1,050,420 of chromosome 19 (GRCh37/hg19 assembly), which is indicated by RepeatMasker to be within a short-interspersed repetitive element of 313 bp.[20] Hence, reliable genotyping or sequencing

primers could not be designed for rs115550680. Imputation of this position in our WES data using the Michigan Imputation Server (imputationserver. sph.umich.edu/index.html) resulted in low-quality imputation for this locus ($r^2 = 0.28$). When we tested the imputed rs115550680 genotypes for their association with AD risk in our cohort, we observed an OR estimate indicative of increased risk for the minor G allele, although this result did not achieve statistical significance (data not shown).

Gene-based analysis of the 20 LOAD GWAS candidate genes using all variants identified in our WES revealed association for variants in *MS4A6A* (n = 5 SNPs, $p = 0.042$) and *ZCWPW1* (n = 11 SNPs, $p = 0.028$) (table e-3). Given that the majority of the missing heritability is thought to reside within low-frequency (5%> MAF >0.5%) and rare

Table 2 AD-risk association analyses of SNPs identified in WES of African American participants

Chr:position (dbSNP142); gene	Maj/min	WES cohort (131 AD vs 107 controls)						Combined cohort (198 AD vs 340 controls)					
		MAF con, %	MAF AD, %	Fisher OR (L95–U95)	Fisher p value	LogReg OR (L95–U95)	LogReg p value	MAF con, %	MAF AD, %	Fisher OR (L95–U95)	Fisher p value	LogReg OR (L95–U95)	LogReg p value
19:1044712 (rs3764647); ABCA7	A/G	17.8	30.2	2 (1.29–3.1)	0.0019	2.53 (1.52–4.21)	0.0004[a]	23.1	29.8	1.41 (1.07–1.87)	0.017[a]	1.47 (1.07–2.02)	0.018[a]
19:1045026 (rs10405305); ABCA7	C/G	4.7	11.5	2.64 (1.26–5.53)	0.0079[a]	3.7 (1.67–8.19)	0.0012[a]	8.3	9.7	1.18 (0.77–1.82)	0.502	1.25 (0.78–2)	0.354
19:1047537 (rs3752239); ABCA7	A/C	0	2.3	NA (NA–NA)	0.0350[a]	—	—	0.4	1.8	4.06 (1.04–15.79)	0.044[a]	—	—
20:55033420 (rs35031530); CASS4	C/T	32.2	24.1	0.67 (0.44–1)	0.0509	0.51 (0.31–0.83)	0.0063[a]	27.8	24.2	0.83 (0.62–1.1)	0.223	0.7 (0.5–0.97)	0.035[a]
11:59950687 (rs114779719); MS4A6A	C/G	4.7	1.2	0.24 (0.06–0.87)	0.0234[a]	0.19 (0.05–0.77)	0.0201[a]	3.5	1.8	0.50 (0.22–1.20)	0.128	0.46 (0.18–1.15)	0.097
7:37907304 (rs10250905); NME8	C/T	18.7	12.6	0.63 (0.38–1.04)	0.0739	0.52 (0.29–0.93)	0.0277[a]	13.7	14.9	1.11 (0.78–1.57)	0.587	0.98 (0.66–1.44)	0.901
7:37923917 (rs62001869); NME8	G/A	6.5	3.8	0.57 (0.25–1.3)	0.2085	0.39 (0.15–0.99)	0.0483[a]	6.3	4.8	0.74 (0.43–1.3)	0.343	0.74 (0.4–1.4)	0.358
7:100014711 (rs6465770); ZCWPW1	T/C	2.8	0.4	0.13 (0.02–1.11)	0.0493[a]	—	—	1.8	1.0	0.57 (0.18–1.77)	0.437	—	—

Abbreviations: AD = Alzheimer disease; con = controls; L95–U95 = upper–lower 95% confidence interval; LogReg = logistic regression; MAF = minor allele frequency; NA = not assessed due to a lack of minor alleles in controls; OR = odds ratio; SNP = single nucleotide polymorphism; WES = whole-exome sequencing.

Results for SNPs that have nominal significance (p < 0.05) either with multivariate logistic regression analysis or with Fisher exact test in the WES cohort and that underwent validation are shown. Results are shown for the WES cohort and for the combined cohort that is composed of the WES and validation cohorts. Variants with total minor allele counts <10 were only assessed by the Fisher exact test.

[a] Significant.

(MAF <0.5%) variants,[21] we also sought to separately assess these specific classes of variants. Evaluation of the low-frequency variants resulted in a stronger AD-risk association with *ZCWPW1* (n = 2 SNPs, p = 0.0099) and with *PTK2B* variants (n = 2 SNPs, p = 0.042). *ABCA7* was the only gene with significant association using rare SNPs (n = 20 SNPs, p = 0.020).

Neuropsychological evaluations were performed in a subset of 319 individuals from our African American cohort (75 AD cases and 244 controls) (table 1). The 8 variants identified by WES that showed nominal AD-risk association were tested for their effects on verbal and visual episodic memory performance (table 3). The 2 *NME8* alleles which associate with lower risk of AD in the WES cohort (rs10250905-T; rs62001869-A), but that did not achieve statistically significant association with AD risk in the combined cohort, demonstrated nominally significant (p < 0.05) or suggestive (0.05 < p < 0.1) association with higher scores for VRDR and IR (VRIR) tests. We also observed a trend for better LMIR scores in *NME8* rs62001869-A carriers. In addition, *ZCWPW1* rs6465770-C, which showed association with decreased risk of AD in the WES cohort but not in the combined cohort, showed nominally significant association with better LMIR and LMDR scores.

DISCUSSION The goal of this study was to assess genes at previously identified LOAD GWAS loci for coding risk variants in African American participants. This is an AD-risk variant discovery study by WES performed in a sizable cohort of 250 African Americans. A prior WES variant discovery study in African Americans was conducted in 7 patients with AD, followed by genotyping of candidate variants in a larger cohort.[7]

In our combined African American cohort of 538 individuals, 2 *ABCA7* variants, rs3764647 and rs3752239, showed nominal significance for association with AD risk. The common variant, rs3764647, has previously been associated with AD risk in an African American AD GWAS.[22] In that study, the *ABCA7* variant rs3764650, which demonstrated significant association with AD risk in a LOAD GWAS conducted in Caucasian cohorts,[12] showed nominally significant association with AD risk (p = 0.019) with an OR estimate (1.27) similar to that in Caucasians, although the most significant *ABCA7* variant was rs3764647 (OR = 1.32, p = 0.0087). Similarly, in our study, we observed only a trend of association with the Caucasian GWAS SNP rs3764650, with an OR estimate (OR = 1.28) almost identical to those in the prior studies[12,22]; and nominal significance with the common missense variant rs3764647 (OR = 1.47, p = 0.018). In our African

Table 3 Memory endophenotype association analyses of SNPs

Chr: position	dbSNP142	Gene	Minor allele count	VRIR			VRDR			VRPR		
				N	β (L95–U95)	p Value	N	β (L95–U95)	p Value	N	β (L95–U95)	p Value
Chr19: 1044712	rs3764647	ABCA7	144	138	0.21 (−1.69 to 2.11)	0.83	137	0.36 (−1.33 to 2.06)	0.67	137	0.14 (−6.61 to 6.89)	0.97
Chr19: 1045026	rs10405305	ABCA7	48	138	−1.12 (−4.34 to 2.1)	0.49	137	0.36 (−2.51 to 3.23)	0.81	137	0.96 (−10.46 to 12.37)	0.87
Chr19: 1047537	rs3752239	ABCA7	2	138	1.03 (−8.3 to 10.35)	0.83	137	0.26 (−8.04 to 8.55)	0.95	137	−4.77 (−37.78 to 28.24)	0.78
Chr20: 55033420	rs35031530	CASS4	169	138	−0.35 (−2.26 to 1.56)	0.72	137	1.63 (−0.05 to 3.31)	0.057	137	5.99 (−0.71 to 12.69)	0.079
Chr11: 59950687	rs114779719	MS4A6A	20	137	9.73 (−3.28 to 22.73)	0.14	136	0.01 (−11.66 to 11.67)	1	136	0.06 (−46.31 to 46.43)	1
Chr7: 37907304	rs10250905	NME8	94	138	1.93 (−0.16 to 4.01)	0.07	137	1.88 (0.02 to 3.75)	0.0477[a]	137	0.43 (−7.1 to 7.95)	0.91
Chr7: 37923917	rs62001869	NME8	39	138	3.04 (0.04 to 6.03)	0.047[a]	137	3.56 (0.93 to 6.2)	0.0084[a]	137	7.71 (−2.98 to 18.39)	0.16
Chr7: 100014711	rs6465770	ZCWPW1	10	138	−2.37 (−8.3 to 3.55)	0.43	137	−2.69 (−7.96 to 2.57)	0.31	137	−7.94 (−28.93 to 13.04)	0.46

Chr: position	dbSNP142	Gene	Minor allele count	LMIR			LMDR			LMPR		
				N	β (L95–U95)	p Value	N	β (L95–U95)	p Value	N	β (L95–U95)	p Value
Chr19: 1044712	rs3764647	ABCA7	144	318	0.73 (−0.48 to 1.95)	0.24	317	0.8 (−0.5 to 2.1)	0.23	317	1.78 (−2.61 to 6.16)	0.43
Chr19: 1045026	rs10405305	ABCA7	48	317	−0.74 (−2.6 to 1.12)	0.43	316	−0.01 (−2 to 1.98)	0.99	316	1.66 (−5.02 to 8.35)	0.62
Chr19: 1047537	rs3752239	ABCA7	2	318	−1.66 (−10.44 to 7.13)	0.71	317	−0.78 (−10.16 to 8.6)	0.87	317	−9.72 (−41.3 to 21.86)	0.55
Chr20: 55033420	rs35031530	CASS4	169	317	−0.58 (−1.73 to 0.58)	0.33	316	−0.67 (−1.91 to 0.56)	0.28	316	−2.66 (−6.82 to 1.5)	0.21
Chr11: 59950687	rs114779719	MS4A6A	20	317	0.73 (−11.62 to 13.07)	0.91	316	−3.41 (−16.6 to 9.78)	0.61	316	−17 (−61.39 to 27.39)	0.45
Chr7: 37907304	rs10250905	NME8	94	318	−0.05 (−1.42 to 1.33)	0.95	317	−0.03 (−1.5 to 1.43)	0.97	317	−0.29 (−5.23 to 4.65)	0.91
Chr7: 37923917	rs62001869	NME8	39	317	2.02 (−0.08 to 4.11)	0.059	316	1.16 (−1.08 to 3.41)	0.31	316	−0.72 (−8.29 to 6.85)	0.85
Chr7: 100014711	rs6465770	ZCWPW1	10	318	4.66 (0.74 to 8.57)	0.0198[a]	317	4.19 (0.01 to 8.38)	0.0497[a]	317	7.28 (−6.89 to 21.45)	0.31

Abbreviations: DR = delayed recall; IR = immediate recall; L95–U95 = upper–lower 95% confidence interval; LM = Logical Memory; PR = percent retention scores; SNP = single nucleotide polymorphism. VR = Visual Reproduction;

Results of association for an additive model where effect size (β) for the minor allele is shown. Memory endophenotypes are from Wechsler Memory Scale Revised.

[a] Significant.

American cohort, these 2 variants are in strong LD ($r^2 = 0.77$), such that the association of rs3764647 is abolished when adjusting for rs3764650 genotypes, indicating that in African Americans, the AD-risk association observed with the Caucasian GWAS SNP rs3764650 may at least in part be explained by the coding variant rs3764647.

This finding is distinct from the observations emerging from *ABCA7* variant discovery efforts in Caucasian cohorts. Two missense coding SNPs, rs3752246[12] and rs4147934,[23] in modest LD with each other were proposed as functional candidate variants accounting for the *ABCA7* locus GWAS signal in Caucasians. Although both of these missense variants were also detected by WES in our cohort, neither had significant AD-risk association in our cohort. More recently, an intronic low-frequency *ABCA7* variant, rs78117248, showed association in a Belgian cohort[24] even after adjustment for the GWAS variants rs3764650, rs4147929, and rs3752246. Rs78117248 was not assessed in our study, as we only focused on coding variants; however, this variant is monomorphic in African American participants according to dbSNP. These results collectively highlight the allelic heterogeneity of the *ABCA7* locus, where different functional risk variants are likely to underlie the GWAS association signal marked by rs3764650, in Caucasian vs African American participants. Our findings support rs3764647 as a functional risk variant underlying this GWAS signal in African American participants. Rs3764647 leads to the amino acid change p.His395Arg in the first extracellular loop of the ABCA7 protein, a region prone to missense mutations. Our study identified AD-risk association with another missense coding SNP, rs3752232 that leads to a p.Thr319Ala change also in the first extracellular loop of ABCA7. While we did not investigate rs3752232 in our larger cohort due to its strong LD with rs3764647 and the slightly better significance of the latter, it remains possible that both of these variants contribute to the risk of AD in African American participants. Although neither rs3764647 nor rs3752232 is predicted to have a deleterious functional consequence by PolyPhen-2 or SIFT,[25,26] they may be contributing to a cumulative AD risk via subtle effects on ABCA7 structure or function in the context of one another, additional *ABCA7* or other genetic variants. Given our findings, both rs3764647 and rs3752232 merit further assessment by downstream functional assays.

We also found evidence of AD-risk association with a rare missense *ABCA7* SNP, rs3752239-C. This variant has an allele frequency of 0.7% in African Americans according to the Exome Variant Server.[19] We observed this variant in <1% of the African American controls and ~2% of the AD cases,

leading to an OR estimate of ~4. AD association with rs3752239 remains significant even after adjusting for the GWAS variant rs3764650, suggesting that this rare variant may confer AD risk in African American participants independently from the common GWAS variant. Furthermore, rs3752239 leads to p. Asn718Thr in the first ABCA7 intracellular domain prior to the first adenosine triphosphate–binding cassette and is predicted to have a possible damaging effect by PolyPhen.[25] These results nominate rs3752239 as a rare, putative functional variant that may confer AD risk in African Americans. Rs3752239 is in strong LD with rs4147918, which leads to a p.Gln1686Arg amino acid change in the second ABCA7 transmembrane domain. Although rs4147918 was not assessed in our larger cohort due to this LD and lack of predicted functional effect, it remains possible that this variant may also contribute to AD risk in African Americans, alone or in combination with rs3752239.

Rs3752239 shows association with AD risk in The International Genomics of Alzheimer's Project (IGAP) Stage I[10] with $p = 0.004$, although the minor allele C has a protective estimate in this study of Caucasian participants, which is opposite to our findings. This variant is more common in Caucasians with MAF = 3.7% in Exome Variant Server.[19] It is possible that the rs3752239 results in the IGAP population reflect LD with other variants rather than a protective functional effect of this variant in Caucasians. Differential functional effects of rs3752239 in African American vs Caucasian populations are also possible.

The *ABCA7* variants that showed significant association in the single-variant analysis in our study had MAFs that fall in the common (MAF >5%) or rare (MAF <0.5%) range. Notably, our gene-based analysis supports the association of multiple rare *ABCA7* SNPs with AD risk in African Americans. In addition, our gene-based analysis identified significant associations with SNPs in *MS4A6A*, *PTK2B*, and *ZCWPW1* that escaped detection by single-variant analysis. For *PTK2B*, significance was observed only with the set of low-frequency SNPs. *ZCWPW1* gene-based analysis shows significance when variants of all frequencies are analyzed, but this result appears to be driven by low-frequency variants in this gene. Hence, there may be additional variants that associate with AD in these genes that escape detection by single-variant analysis approaches.

In this study, we also aimed to assess whether coding variants in the LOAD candidate genes modulate cognitive endophenotypes. We hypothesized that some genetic variants that associate with AD risk in the WES cohort would also influence memory endophenotypes. We observed that the minor alleles of

rs6465770 in *ZCWPW1*, as well as rs10250905 and rs62001869 in *NME8* that associate with lower AD risk in the WES cohort also have nominally significant associations with better memory scores. Although these variants did not retain significant AD-risk associations in the larger cohort, the biologically concordant association of their minor alleles with better verbal (rs6465770) or visual (rs10250905 and rs62001869) memory scores is noteworthy. Furthermore, rs6465770 in *ZCWPW1* is one of the 2 low-frequency variants that show association in the gene-based analysis. The difference between the memory and AD-risk association results could arise from false-positive or false-negative findings in one of the analyses; or an inherent difference between the overall cohort vs the memory endophenotype subset cohort, although the latter was selected only based on the availability of cognitive results. These findings, which require replication in larger cohorts and longitudinal studies, may imply that these variants may have an influence on the progression of memory decline.

Although informative, this study has limitations. Despite being the largest to date variant discovery effort by WES in African Americans, both our WES and entire cohort have limited power to detect AD-risk associations. None of the variants highlighted in this study would remain significant after corrections for multiple testing for 193 variants in the WES and 8 selected variants in the combined cohort. Likewise, the cohort with cognitive endophenotypes was limited in size, and the cognitive endophenotype associations would not withstand corrections for multiple testing. Given this, the findings from our study require replication in larger independent cohorts. Based on the MAF and OR estimates in our combined cohort, detection of AD-risk association for the rare missense SNP rs3752239 with 80% power at $\alpha = 0.05$ requires a sample of at least 1,124 AD vs 1,124 controls.

Our study represents a systematic variant discovery conducted by WES in African Americans to identify coding AD-risk variants within genes at the 20 LOAD-risk GWAS loci. In our WES cohort of 238 subjects, of the variants that lead to a coding change, most were in *ABCA7*, and the only variants that remained nominally significant in the combined cohort of 538 participants were *ABCA7* missense variants. Furthermore, the rare *ABCA7* SNPs in our study had significant gene-based association. These findings are similar to those observed in Caucasians[23,27,28] and highlight the unique tendency of ABCA7 to harbor coding AD-risk variants.

The significant gene-based associations observed for *MS4A6A*, *PTK2B*, and *ZCWPW1* suggest that there may also be coding AD-risk variants in these genes. Lack of single-variant or gene-based associations with coding variants in the other AD GWAS genes in our study may be multifactorial including insufficient power. It is also possible that risk variants for many of the LOAD genes primarily reside in functional, noncoding regions.[29,30] Finally, given that *ABCA7* is the only Caucasian AD GWAS locus replicated at genome-wide significance in African Americans,[8] lack of identification of coding variants in the other GWAS genes may be due to their lack of influence in this population. This possibility seems unlikely because other Caucasian LOAD GWAS loci, namely *CR1*, *BIN1*, *EPHA1*, and *CD33*, also have evidence of association in African Americans albeit below genome-wide significance.[8]

Our findings also highlight the presence of allelic heterogeneity between different ethnic groups and underscore the importance of variant discovery efforts in both Caucasian and non-Caucasian participants. Our study provides support for the enrichment of *ABCA7* with coding AD-risk variants, suggests that *MS4A6A*, *PTK2B*, and *ZCWPW1* may also harbor AD-risk variants, nominates coding variants in *ZCWPW1* and *NME8* that may be protective by modulating cognition, and importantly reveals a list of coding variants in LOAD candidate genes that can guide future variant discovery and replication efforts in this understudied population.

REFERENCES

1. Tang MX, Cross P, Andrews H, et al. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. Neurology 2001;56:49–56.
2. 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.
3. Plassman BL, Langa KM, Fisher GG, et al. Prevalence of dementia in the United States: the aging, demographics, and memory study. Neuroepidemiology 2007;29:125–132.
4. Williams MM, Meisel MM, Williams J, Morris JC. An interdisciplinary outreach model of African American recruitment for Alzheimer's disease research. Gerontologist 2011;51(suppl 1):S134–S141.
5. Williams MM, Scharff DP, Mathews KJ, et al. Barriers and facilitators of African American participation in Alzheimer's disease biomarker research. Alzheimer Dis Assoc Disord 2010;24(suppl):S24.
6. Darnell KR, McGuire C, Danner DD. African American participation in Alzheimer's disease research that includes brain donation. Am J Alzheimers Dis Other Demen 2011; 26:469–476.
7. Logue MW, Schu M, Vardarajan BN, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. Alzheimers Dement 2014;10:609–618.e11.
8. Reitz C, Jun G, Naj A, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ε4, and the risk of late-onset Alzheimer disease in African Americans. JAMA 2013;309:1483–1492.
9. Jin SC, Carrasquillo MM, Benitez BA, et al. TREM2 is associated with increased risk for Alzheimer's disease in African Americans. Mol Neurodegener 2015;10:19.
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. Chen R, Im H, Snyder M. Whole-exome enrichment with the Agilent SureSelect Human All Exon platform. Cold Spring Harb Protoc 2015;2015:626–633.
12. 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.
13. Wechsler D. WMS-R: Wechsler Memory Scale-Revised. San Antonio, TX: Psychological Corporation; 1987.
14. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996; 49:1373–1379.
15. 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.
16. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–265.
17. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 2012;91:224–237.
18. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. ISBN 3-900051-07-0, 2014.
19. Exome Variant Server. NHLBI GO Exome Sequencing Project (ESP) [online]. Seattle, WA: Exome Variant Server. Available at: evs.gs.washington.edu/EVS/. Accessed in August 2016.
20. RepeatMasker Open-4.0 [computer program]. 2013–2015. Available at: repeatmasker.org. Accessed in April 2015.
21. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet 2014;95:5–23.
22. Logue MW, Schu M, Vardarajan BN, et al. A comprehensive genetic association study of Alzheimer disease in African Americans. Arch Neurol 2011;68:1569–1579.
23. Holton P, Ryten M, Nalls M, et al. Initial assessment of the pathogenic mechanisms of the recently identified Alzheimer risk Loci. Ann Hum Genet 2013;77:85–105.
24. 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.
25. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–249.
26. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 2012;40: W452–W457.
27. Vardarajan BN, Ghani M, Kahn A, et al. Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci. Ann Neurol 2015; 78:487–498.
28. 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.
29. Allen M, Zou F, Chai HS, et al. Novel late-onset Alzheimer disease loci variants associate with brain gene expression. Neurology 2012;79:221–228.
30. Allen M, Kachadoorian M, Carrasquillo MM, et al. Late-onset Alzheimer disease risk variants mark brain regulatory loci. Neurol Genet 2015;1:e15. doi: 10.1212/NXG. 0000000000000015.

Neurology® Genetics

African American exome sequencing identifies potential risk variants at Alzheimer disease loci

Aurelie N'Songo, Minerva M. Carrasquillo, Xue Wang, et al.

This information is current as of April 7, 2017