Population-based Analysis of Alzheimer's Disease Risk Alleles Implicates Genetic Interactions

Mark T.W. Ebbert, Perry G. Ridge, Andrew R. Wilson, Aaron R. Sharp, Matthew Bailey, Maria C. Norton, JoAnn T. Tschanz, Ronald G. Munger, Christopher D. Corcoran, and John S.K. Kauwe

Background: Reported odds ratios and population attributable fractions (PAF) for late-onset Alzheimer's disease (LOAD) risk loci (*BIN1*, *ABCA7*, *CR1*, *MS4A4E*, *CD2AP*, *PICALM*, *MS4A6A*, *CD33*, and *CLU)* come from clinically ascertained samples. Little is known about the combined PAF for these LOAD risk alleles and the utility of these combined markers for case–control prediction. Here we evaluate these loci in a large population-based sample to estimate PAF and explore the effects of additive and nonadditive interactions on LOAD status prediction performance.

Methods: 2419 samples from the Cache County Memory Study were genotyped for *APOE* and nine LOAD risk loci from AlzGene.org. We used logistic regression and receiver operator characteristic analysis to assess the LOAD status prediction performance of these loci using additive and nonadditive models and compared odds ratios and PAFs between AlzGene.org and Cache County.

Results: Odds ratios were comparable between Cache County and AlzGene.org when identical single nucleotide polymorphisms were genotyped. PAFs from AlzGene.org ranged from 2.25% to 37%; those from Cache County ranged from .05% to 20%. Including non-*APOE* alleles significantly improved LOAD status prediction performance (area under the curve = .80) over *APOE* alone (area under the curve = .78) when not constrained to an additive relationship ($p < .03$). We identified potential allelic interactions (p values uncorrected): *CD33-MS4A4E* (synergy factor = 5.31; $p < .003$) and *CLU-MS4A4E* (synergy factor = 3.81; $p < .016$).

Conclusions: Although nonadditive interactions between loci significantly improve diagnostic ability, the improvement does not reach the desired sensitivity or specificity for clinical use. Nevertheless, these results suggest that understanding gene–gene interactions may be important in resolving Alzheimer's disease etiology.

Key Words: Alzheimer's disease, epistasis, genetic interactions, population attributable fraction, odds ratio, risk

Researchers have implicated several genes associated with late-onset Alzheimer's disease (LOAD) including *APOE*. *APOE ε4* increases LOAD risk and *APOE ε2* reduces risk (1–4). According to AlzGene.org (5), nine additional genes significantly affect LOAD risk; *BIN1* (rs744373), *ABCA7* (rs3764650), *CR1* (rs3818361), *MS4A4E* (rs670139), and *CD2AP* (rs9349407) are associated with increased risk for LOAD, and *PICALM* (rs3851179), *MS4A6A* (rs610932), *CD33* (rs3865444), and *CLU* (rs11136000) are associated with decreased risk (6–10). Only one study to date has examined the contribution of these nine risk alleles to LOAD status prediction (11). Verhaaren *et al*. (11) calculated an additive genetic risk score and compared LOAD status prediction performance of age, gender, and *APOE ε4* genotype using logistic regression with and without the additive genetic risk score. The genetic risk score did not improve prediction performance significantly, suggesting that the nine alleles may not be diagnostically useful when constrained to an additive relationship. The assumption of additive relationships between risk loci is common but is likely to be an oversimplification of the underlying biology for LOAD and other complex diseases (12–14). In fact, there may be underlying gene–gene interactions not examined in the Verhaaren *et al*. (11) study or others that improve LOAD status prediction performance.

Some of the population attributable fractions for these nine loci have been reported individually and in different combinations (6,8,9); however, no study to date has reported the combined population attributable fraction for all nine risk alleles. Furthermore, previously reported odds ratios and population attributable fractions are from clinically ascertained samples rather than a population-based sample (6–10). The latter may provide a more reliable measure of population risk because clinically ascertained samples select for disease, enriching risk alleles in the sample.

In this study, we estimated the allelic odds ratios and population-attributable fractions for *APOE ε2*, *APOE ε4*, and the nine non-*APOE* LOAD risk alleles in a large population-based sample. We also extended the genetic risk score used by Verhaaren *et al*. (11) by testing whether the nine non-*APOE* alleles contribute significantly to LOAD status prediction when interactions between loci are not constrained to additive relationships.

Methods and Materials

Sample Collection

The Cache County Study on Memory Health and Aging was initiated in 1994 (15). This cohort of 5092 individuals represented approximately 90% of the Cache County population aged 65 and older. Specific details about data collection, obtaining consent, and phenotyping individuals in the Cache County population have been reported previously (15). Briefly, case-control status was determined in four triennial waves of data collection in a

From the Department of Biology (MTWE, PGR, ARS, MB, JSKK), Brigham Young University, Provo; ARUP Institute for Clinical and Experimental Pathology (MTWE, PGR, ARW), and Departments of Family Consumer and Human Development (MCN), Psychology (JTT), Nutrition, Dietetics, and Food Sciences (RGM), and Mathematics and Statistics (CDC) and Center for Epidemiologic Studies (MCN, JTT, RGM, CDC) Utah State University, Logan, Utah.

Address correspondence to John S. K. Kauwe, Ph.D., 675 WIDB, Provo, UT 84602; E-mail: kauwe@byu.edu.

Received Mar 7, 2013; revised Jun 24, 2013; accepted Jul 2, 2013.

0006-3223/$36.00
http://dx.doi.org/10.1016/j.biopsych.2013.07.008

multistage dementia screening and assessment protocol. The first stage of screening consisted of administration of the Modified Mini-Mental State Exam-Revised (16). Screen-positive individuals and a randomly selected 19% designated subsample were invited to complete subsequent stages of evaluation consisting of an informant interview and the next stage, a clinical assessment including neuropsychological testing. The clinical assessment results were reviewed by a geropsychiatrist and neuropsychologist and preliminary diagnoses of dementia or other cognitive disorders were assigned. Those carrying a diagnosis of dementia or its prodrome were invited to complete standard laboratory tests for dementia, a magnetic resonance scan, and a geropsychiatrist examination. Final case–control status was determined by an expert panel of clinicians including study geropsychiatrists, neuropsychologists, a neurologist, and cognitive neuroscientist. Diagnoses of AD followed National Institute of Neurological Disorders and Stroke–Alzheimer's Disease and Related Disorders Association criteria (17), and cases included possible or probable AD. Control subjects were identified as those who were diagnosed with no dementia (per clinical assessment) or whose cognitive test result was negative at each preceding screening stage. Persons with incomplete screening results (i.e., those who were screen positive at one stage but did not complete the subsequent stage) or missing genotype data were excluded from the analyses, leaving 2093 participants without dementia (control subjects) and 326 persons with LOAD (cases). All study procedures were approved by the Institutional Review Boards of Utah State, Duke University, and Johns Hopkins University.

DNA from the 2419 Cache County study participants was genotyped for the nine non-*APOE* LOAD risk alleles in the AlzGene.org "AlzGene Top Results" list (5) using TaqMan Assays (Table 1). Genotyping failed for rs3764650 (*ABCA7*) and rs3818361 (*CR1*), so we selected rs3752246 and rs6656401 to represent the effects reported by *ABCA7* and *CR1* for AD risk, respectively. The *CR1* single nucleotide polymorphisms (SNPs) are in high linkage disequilibrium ($D' = .995$, $R^2 = .84$), and both *ABCA7* SNPs are within 10 kilobases of each other and rs3752246 was reported as significant by Naj *et al.* (9). *APOE ε2* and *APOE ε4* were previously genotyped as part of the Cache County study (15).

Statistical Analyses

All statistical analyses were performed in R (18). We used logistic regression and receiver operating characteristic curve analysis to assess case–control predictive performance of the nine non-*APOE* alleles. Specifically, we tested whether the non-*APOE* alleles significantly improved LOAD status prediction performance over models excluding the non-*APOE* alleles. Two types of models were generated: additive risk profiles and genotype models to test potential additive and nonadditive relationships, respectively. To assess efficacy of each model, we measured LOAD status prediction performance using the area under the curve (AUC) of the receiver operating characteristic curves. All models were adjusted for age and gender. A separate model using only age and gender was also generated to establish reference values.

We calculated three additive risk scores for participants in the Cache County Study to measure LOAD status prediction performance for the nine non-*APOE* LOAD risk alleles. Specifically, the following risk profiles were calculated: 1) *APOE* alone; 2) the nine LOAD risk alleles with *APOE*; and 3) the nine LOAD risk alleles without *APOE*. The risk allele (whether the major or the minor allele) and associated beta coefficient were used for each locus. We calculated additive risk scores as the sum of the risk across all alleles (Equation 1):

$$RP = \sum_{i}^{n} \beta_i N_i,$$

Table 1. Summary Statistics for Significant Markers

SNP	Nearest Gene	MAF		Odds Ratio		PAF	
		AlzGene	Cache Co.	AlzGene (95% CI)	Cache Co. (95% CI)	AlzGene	Cache Co.
rs3752246[a]	*ABCA7*	.10	.18	1.23 (1.18–1.28)	.94 (.76–1.17)	2.25	4.65
rs7412	*APOE2*	.06	.09	.62 (.46–.85)	.89 (.63–1.22)	36	10
rs429358	*APOE4*	.22	.17	3.68 (3.30–4.11)	2.51 (2.07–3.04)	37	20
rs744373	*BIN1*	.29	.30	1.17 (1.13–1.20)	1.02 (.85–1.22)	4.61	.54
rs9349407	*CD2AP*	.29	.28	1.12 (1.08–1.16)	1.03 (.85–1.23)	3.29	.70
rs3865444	*CD33*	.31	.34	.89 (.86–.92)	1.00 (.84–1.19)	7.63	.05
rs11136000	*CLU*	.38	.39	.88 (.86–.91)	.88 (.74–1.04)	7.85	7.98
rs6656401	*CR1*	.19	.19	1.19 (1.09–1.30)	.92 (.74–1.13)	3.49	6.84
rs670139	*MS4A4E*	.41	.41	1.08 (1.05–1.11)	1.0 (.84–1.18)	3.14	.05
rs610932	*MS4A6A*	.42	.43	.90 (.88–.93)	.89 (.76–1.06)	5.81	6.33
rs3851179	*PICALM*	.35	.38	.88 (.86–.91)	.85 (.72–1.01)	8.19	9.69
				Combined PAF (All Alleles)		75	51
				Combined PAF (Excluding APOE)		38	32

MAFs, odds ratios, and PAFs were calculated for all single nucleotide polymorphisms using both data from AlzGene.org and the Cache County (Co.) population-based study. PAFs are reported as percentages. For better interpretation and comparison to previous studies, the risk allele for each locus (whether the major or the minor allele) was used to calculate PAFs, but the minor allele was used for odds ratios. MAFs are comparable between AlzGene.org and the Cache County data. Odds ratios are generally similar except that *ABCA7* and *CR1* differ in direction. Individual PAFs in Cache County varied in magnitude compared with those calculated for AlzGene.org. Combined population attributable fractions were also lower in Cache County. As expected, *APOE ε4* and *APOE ε2* have strong population effects, whereas the remaining alleles have minimal individual effect. On the basis of the AlzGene.org data, combined PAFs suggest the combined effect of the nine non-*APOE* alleles is approximately equal to *APOE ε2* or *APOE ε4* alone; however, the nine non-*APOE* alleles appear to have a larger effect than either *APOE* allele in the Cache County data.

CI, confidence interval; MAF, minor allele frequency; PAF, population attributable fraction.

[a]The single nucleotide polymorphisms (SNPs) for *ABCA7* (rs3752246) was not reported on AlzGene.org, but was reported in Naj *et al.* (9) as significant and was used in place of rs3764350.