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.

where β equals a previously calculated risk allele beta coefficient from odds ratios (β = ln(odds ratio)) reported by AlzGene.org (accessed February 2012), and *N* equals the subject's number of risk alleles. *APOE ε2* and *APOE ε4* were coded jointly into a single class variable as 22, 23, 24, 33, 34, and 44.

We also tested genotype models using genotype data in place of the risk profile score. We generated the following genotype models: 1) *APOE* alone; 2) the nine LOAD risk alleles with *APOE*; and 3) an optimized model. Using genotypes does not constrain the model to an additive relationship, allowing for other genetic models within each locus. The optimized model was generated using a stepwise regression method to test if interactions between loci contribute to LOAD status prediction and was selected using Akaike's information criterion. To test for and avoid overfitting, we included three random variables while generating the optimized model. These variables were generated randomly with respect to all genotype and phenotype data in our study and were included to provide evidence that the selected variables provide meaningful information (19). Although the absence of all random variables in the model does not guarantee the model was not overfit, it does suggest the included variables provide useful diagnostic information.

Synergy factors, statistics that measure the strength of allelic interactions in case–control studies (13,20), were calculated for any statistically significant allelic interactions using logistic regression. All synergy factors were adjusted for age, gender, and *APOE ε4* by including only the main effects of the interacting alleles, the interaction term between the alleles, age, gender, and the number *APOE ε4* alleles (Status = allele1*allele2 + age + gender + *APOE*4num). Synergy factor confidence intervals (CIs) were calculated using the interaction term coefficient ± 1.96 * standard error of the parameter estimate of the interaction term.

Odds ratios and population attributable fractions were also calculated. Odds ratios here estimate the relative risk of Alzheimer's disease given allelic exposure, and population attributable fractions estimate the proportional decrease in LOAD cases that would occur if the risk factor were removed from the population. Odds ratios were calculated only for the Cache County subjects, but population attributable fractions were calculated for both Cache County subjects and the pooled AlzGene samples using published odds ratios and minor allele frequencies from AlzGene.org. We calculated population attributable fractions using Equation 2 (9,21):

$$PAF = \frac{p(OR-1)}{p(OR-1)+1},$$

in which *p* equals the allele frequency, and OR is the odds ratio. A combined population attributable fraction was calculated for all risk factors and just the nine non-*APOE* risk factors using Equation 3 (9,21,22) to estimate the proportional decrease in LOAD cases if all included risk factors were removed from the population:

$$cPAF = 1 - \prod_{j=1}^{n}(1-PAF_j).$$

In this equation, PAF_j represents previously calculated PAFs from Equation 2, and *n* is the number of loci included in the combined PAF. 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 population attributable fractions, but the minor allele was used for odds ratios.

Results

Sample Demographics

The sample consisted of 1406 females and 1013 males. The mean age and standard deviation were 75.13 and 7.29 years, respectively. Mean age was significantly different between cases and controls ($p < 2.2e-16$), as were the proportion of males in each group ($p < .04$; Table S1 in Supplement 1). Similarly, mean age was significantly different between participants included in the study and those excluded for reasons previously mentioned ($p < 2.2e-16$; Table S2 in Supplement 1). The proportion of male subjects, however, was not significantly different between included and excluded participants ($p < .29$).

Odds Ratios

Odds ratios calculated for the Cache County data were generally comparable in direction and magnitude to odds ratios from AlzGene.org when identical SNPs were genotyped. *ABCA7* and *CR1* varied, but a different SNP was genotyped for *ABCA7*, and the 95% CIs for *CR1* overlap between AlzGene.org and Cache County results (Table 1). Odds ratios from meta-analyses on AlzGene.org for *ABCA7* and *CR1* are 1.23 (95% CI 1.18–1.28) and 1.19 (95% CI 1.09–1.30), respectively, whereas the data from Cache County were .94 (95% CI .76–1.17) and .92 (95% CI .74–1.13), respectively. No alleles deviated significantly from Hardy-Weinberg equilibrium.

Population Attributable Fraction. Population attributable fractions as calculated from AlzGene.org data ranged from 2.25% to 37%, whereas those from Cache County ranged from .05% to 20% (Table 1). The highest risks were attributed to *APOE ε4* (AlzGene = 37%; Cache = 20%) and lack of the *APOE ε2* (AlzGene = 36%; Cache = 10%), whereas the next highest risk was attributed to *PICALM* (AlzGene = 8.19%; Cache = 9.69%). The smallest risk for AlzGene.org was from *ABCA7* (2.2%), and the smallest for the Cache County data were *CD33* and *MS4A4E* (.05%). Combined population attributable fractions for all LOAD risk alleles (including *APOE*) were 75% and 51% for AlzGene.org and Cache County, respectively. Using only the nine non-*APOE* alleles, combined population attributable fractions were 38% and 32% for AlzGene.org and Cache County, respectively.

LOAD Status Prediction Performance. The non-*APOE* alleles combined with *APOE* (AUC = .782) did not improve LOAD status prediction performance over *APOE* alone (AUC = .783) when constrained to an additive model (Figure S1 in Supplement 1), as previously reported (11), nor did the non-*APOE* alleles without *APOE* (AUC = .728) significantly improve LOAD status prediction performance over age and gender alone (AUC = .727; $p < .2372$). The model using all genotype data (full genotype model), when not constrained to an additive relationship (AUC = .796) did, however, improve LOAD status prediction performance significantly over *APOE* alone (AUC = .783; $p < .03$; Figure 1). Moreover, the optimized model allowing for interactions between loci (AUC = .82) improved significantly over the full genotype model ($p < 8.39e-07$). All three genotype models improve prediction performance significantly over age and gender alone. None of the random variables previously mentioned were selected for the optimized model. Selected variables and interactions for the optimized model are as follows: rs3752246, rs6656401, rs11136000, rs610932, rs3865444, rs670139, Age, APOE, rs3865444:rs670139, rs11136000:rs670139, rs3752246:APOE, rs3752246:rs610932, and rs670139:Age.

Locus Interactions. Investigating the optimized genotype model revealed two statistically significant alleles and two significant allelic interactions, although the *p* values were not

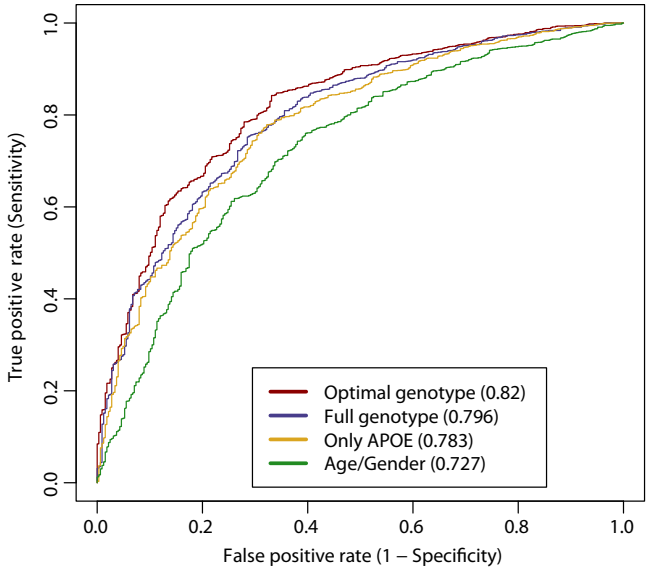

Figure 1. Non-*APOE* late-onset Alzheimer's disease (LOAD) risk loci contributions to LOAD status prediction performance. Three logistic regression models based on age, gender, and genetic information for *APOE* and the non-*APOE* LOAD risk loci illustrate the contribution of the non-*APOE* LOAD risk loci in LOAD status prediction performance. The models are as follows: *APOE* alone (Only *APOE*), all loci (Full genotype), and the optimized model (Optimal genotype). A fourth model using only age and gender (Age/Gender) was also generated as a baseline. The optimized model was optimized using Akaike's Information Criterion. Comparing the full genotype model to *APOE* alone demonstrates that the LOAD risk loci contribute significantly to LOAD status prediction performance ($p < .03$), and the optimized model improves significantly over the full genotype model ($p < 8.39e-07$). Area under the curve is listed in parentheses within the legend.

corrected for multiple testing. Genotypes A/G ($p < .02$) and G/G ($p < .03$) in rs6656401 (*CR1*) were significant individually. The significant interactions were between the rs3865444 C/C (*CLU*) genotype and the rs670139 G/G (*MS4A4E*) genotype ($p < .016$; synergy factor 3.81, 95% CI 1.28–11.32) and the rs11136000 C/C (*CD33*) genotype and the rs670139 G/G (*MS4A4E*) genotype ($p < .003$; synergy factor 5.31, 95% CI 1.79–15.77).

Discussion

Recent research has identified several alleles that may prove useful in resolving Alzheimer's etiology (6–10), but until now there had not been an assessment of their population attributable fraction in a large, population-based sample. Similarly, deeper interrogation of the diagnostic utility of the Alzheimer's disease candidate genes is needed. Verhaaren *et al.* (11) explored the diagnostic utility based on an additive relationship, which we replicated in this work, but they did not test locus interactions—a major aim of this research. During this process, we also estimated allelic odds ratios and population attributable fractions.

The data reported in this study are generalizable to other U.S. populations of northern European descent. The Cache County population has been included in the Centre d'Etude du Polymorphisme Humain families that are used to represent the European sample in the HapMap project (23,24). Utah's early pioneers were mostly unrelated and originated from various European locations (25–27), which is necessary for generalizability. The AlgGene.org data, a meta-analysis, varies between loci but is largely Caucasian-based as well. Many of the loci include populations of African, Asian, and Hispanic descent, but the sample sizes for these populations are much smaller than the Caucasian populations.

Odds Ratios

We compared Cache County odds ratios to those reported in the meta-analyses on AlzGene.org and found them comparable. Minor differences were observed in *ABCA7* and *CR1* where we genotyped SNPs that are not listed on AlzGene.org. Specifically, minor alleles for both *ABCA7* and *CR1* were considered risk alleles (odds ratio >1) according to data on AlzGene.org, whereas odds ratios in the Cache County data suggest decreased risk, although the CIs from both studies are broad and overlap each other, so they may not be significantly different. Possible causes include that 1) differences in sample ascertainment between clinical and population studies (e.g., the cases in clinically ascertained samples are generally younger than those in the Cache County sample; see AlzGene.org, Table S1 in Supplement 1); and 2) allelic odds ratios are not adjusted for age, gender, and other loci, nor are they adjusted for undiscovered or uncharacterized allelic interactions (13,14,28,29).

Clinical and population studies differ in sample ascertainment. Clinically ascertained cases and controls are selected to minimize confounding variables and maximize contrast between the true underlying causes by minimizing known differences between the two groups except for the phenotype of interest. Population-based studies, however, are designed to represent true population characteristics such as allele frequencies, odds ratios, and population attributable fractions, as reported here. Although population-based studies are ideal for reporting population characteristics, small sample sizes for cases may reduce the accuracy of odds ratio and PAF estimates. Because of the natural differences between clinically ascertained case-control and population-based studies, it is important to leverage the strengths of each of them.

The complex nature of Alzheimer's disease inheritance, however, suggests that variations between studies may exist because allelic odds ratios are not adjusted for age, gender, and other loci, nor are they adjusted for undiscovered and uncharacterized allelic interactions. Each of these factors plays a significant role in Alzheimer's etiology, and not adjusting for them introduces error into odds ratio estimates. Allelic interactions also likely contribute to the "missing heritability" in Alzheimer's disease. No single genetic locus characterizes Alzheimer's etiology. *APOE* alone is highly predictive, but the genetic loci included here also appear to influence Alzheimer's susceptibility, as reported in this study and others (6–10). Furthermore the effects of *APOE* vary between ethnic groups (30–34). Failure to replicate established genome-wide association study findings in some populations (13,35) further suggests the possible influence of environmental factors and gene–environment and gene–gene interactions.

Population Attributable Fractions

Cache County population attributable fractions varied in magnitude when compared with those calculated from AlzGene.org data. Combined population attributable fractions were lower in Cache County. As expected *APOE ε4* and *APOE ε2* have strong population effects, whereas the remaining alleles have minimal individual effects. On the basis of AlzGene.org data, combined population attributable fractions suggest the combined effect of the nine non-*APOE* alleles is approximately equal to *APOE ε2* or *APOE ε4* alone; however, the combined non-*APOE* alleles appear to have a larger effect than either *APOE* allele in the

Cache County data. The Cache County data are of value because they are population-based and better represent risks within populations—the purpose of the PAF statistic. Despite being more conservative than other estimates (combined), however, the population attributable fractions reported in this study may still be inflated because they are based on the unadjusted allelic odds ratios and because the exposure frequency for the genotyped SNPs may vary from the functional variants they represent. They may also be biased as a consequence of potentially inaccurate odds ratios due to the small sample size for cases, as mentioned earlier. Future estimates are also likely to change as allelic interactions are discovered and incorporated into the calculations.

Diagnostic Utility

Verhaaren *et al*. (11) demonstrated that the nine non-*APOE* genes do not improve LOAD status prediction performance when constrained to an additive relationship, which we confirmed in this study. When unconstrained, however, the top nine alleles improved LOAD status prediction performance significantly, demonstrating that these alleles may provide more information as we better understand their epistatic relationships. The optimized model further improved LOAD status prediction performance and revealed *CLU-MS4A4E* and *CD33-MS4A4E* interactions that may prove valuable in Alzheimer's research. Synergy factors for both interactions suggest that being homozygous for both alleles in either interaction increases risk. Yet although these data suggest the additional LOAD risk alleles significantly improve LOAD status prediction performance, the improvement is marginal and does not reach the desired sensitivity or specificity for clinical use.

The optimized model clearly improves LOAD status prediction performance over the full genotype model and over *APOE* alone, suggesting allelic interactions may be useful for diagnostic purposes; however, the *p* values were not corrected for multiple testing. As such, these interactions need to be tested in an independent data set. It is also possible the optimized model is overfit; however, the random variables included in the model selection process were not selected for the final model, lending evidence that the final variables included provide nonrandom information. The revealed interactions also have strong synergy factors, suggesting they may be important. Furthermore, the genotype model with all alleles improves LOAD status prediction performance over *APOE* alone, lending support for underlying relationships amongst the factors included in the model.

Implications and Future Directions

The results presented here offer evidence that gene–gene interactions play a role in Alzheimer's susceptibility; however, the reported interactions do not appear to improve LOAD status prediction performance by an amount that is relevant in a clinical diagnostic setting. These results do suggest that to fully understand the genetic basis of Alzheimer's disease risk we must improve our efforts to characterize gene–gene and gene–environment interactions.

Additionally, environmental factors have not received as much attention as genetic factors in Alzheimer's research and should be thoroughly investigated (12). Although the *CLU-MS4A4E* and *CD33-MS4A4E* interactions appear to have strong effects in the Cache County study, there may be unmeasured environmental factors that increase the effect of these interactions in the Cache County population. Other research has shown that only 30% of Alzheimer's disease is explained by known genes, demonstrating that environmental effects and gene by environment interactions will be essential in future studies (36).

The *CLU-MS4A4E* and *CD33-MS4A4E* interactions have not been previously reported leaving the biological foundation in question. Using IPA (Ingenuity Systems, Redwood City, California; www.ingenuity.com), we explored possible interactions between each pair and found that, although no information is available for *MS4A4E* specifically, both *CLU* and *CD33* interact indirectly with *MS4A2* (Figure S2 in Supplement 1). According to IPA, both thioacetamide and *TGFB1* act indirectly on both *CLU* and *MS4A2* (Figure S2A in Supplement 1). *CLU* also binds to *BCL2L1*, which is acted on by *MS4A2*. Likewise, *CD33* acts on *PTPN6,* which binds to *MS4A2* and *CD33* binds to *CBL*, which then acts on *MS4A2* (Figure S2B in Supplement 1). Both *MS4A4E* and *MS4A2* are members of the membrane-spanning four-domain gene family. A complete IPA legend is available in Ingenuity's website (http://ingenuity.force.com/ipa/articles/Feature_Description/Legend).

Overall, the results presented in this article suggest that gene–gene interactions (epistasis) may play an important role in Alzheimer's etiology. Although discovering and characterizing epistatic interactions is a nontrivial task, researchers and consortiums must heed the plentiful evidence that Alzheimer's is driven by complex gene–gene and gene–environment interactions.

This work was supported by the National Institutes of Health (Grant Nos. R01AG11380, R01AG21136, R01AG31272, and R01AG042611), the Alzheimer's Association (Grant No. MNIRG-11-205368), the Utah Science, Technology, and Research initiative, the Utah State University Agricultural Experiment Station, and the Brigham Young University Gerontology Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

We thank the participants and staff of the Dementia Progression Study, the Utah Population Database, and the Cache County Study on Memory Health and Aging for their important contributions to this work. Additionally, we acknowledge the assistance of Drs. David Ward and Ned Weinshenker.

The authors report no biomedical financial interests or potential conflicts of interest.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2013.07.008.

1. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, *et al*. (1993): Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261: 921–923.
2. Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, *et al*. (1993): Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 43:1467–1472.
3. St Clair D, Rennie M, Slorach E, Norrman J, Yates C, Carothers A (1995): Apolipoprotein E epsilon 4 allele is a risk factor for familial and sporadic presenile Alzheimer's disease in both homozygote and heterozygote carriers. *J Med Genet* 32:642–644.
4. Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, *et al*. (1993): Apolipoprotein E: High-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A* 90: 1977–1981.
5. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE (2007): Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nat Genet* 39:17–23.
6. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, *et al*. (2009): Genome-wide association study identifies variants at CLU

and PICALM associated with Alzheimer's disease. *Nat Genet* 41: 1088–1093.
7. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, *et al*. (2011): Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 43:429–435.
8. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, *et al*. (2009): Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 41:1094–1099.
9. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, *et al*. (2011): Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 43:436–441.
10. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, *et al*. (2010): Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 303:1832–1840.
11. Verhaaren BF, Vernooij MW, Koudstaal PJ, Uitterlinden AG, Duijn CM, Hofman A, *et al*. (2013): Alzheimer's disease genes and cognition in the nondemented general population. *Biol Psychiatry* 73:429–434.
12. Bullock JM, Medway C, Cortina-Borja M, Turton JC, Prince JA, Ibrahim-Verbaas CA, *et al*. (2013): Discovery by the Epistasis Project of an epistatic interaction between the GSTM3 gene and the HHEX/IDE/KIF11 locus in the risk of Alzheimer's disease. *Neurobiol Aging* 34: 1309.e1–1309.e7.
13. Combarros O, Cortina-Borja M, Smith AD, Lehmann DJ (2009): Epistasis in sporadic Alzheimer's disease. *Neurobiol Aging* 30:1333–1349.
14. Moore JH, Williams SM (2005): Traversing the conceptual divide between biological and statistical epistasis: Systems biology and a more modern synthesis. *BioEssays* 27:637–646.
15. Breitner JC, Wyse BW, Anthony JC, Welsh-Bohmer KA, Steffens DC, Norton MC, *et al*. (1999): APOE-epsilon4 count predicts age when prevalence of AD increases, then declines: The Cache County Study. *Neurology* 53:321–331.
16. Tschanz JT, Welsh-Bohmer KA, Plassman BL, Norton MC, Wyse BW, Breitner JC, *et al*. (2002): An adaptation of the modified Mini-Mental State Examination: analysis of demographic influences and normative data: The Cache County Study. *Neuropsychiatry Neuropsychol Behav Neurol* 15:28–38.
17. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984): Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34:939–944.
18. R Core Team (2012): *R: A Language and Environment for Statistical Computing. 2.15.1 Ed*. Vienna, Austria: R Foundation for Statistical Computing.
19. Gonzalez NA, Swain NR, Obregon O, Buehler BD, Williams GP, Nelson EJ, *et al*. (2012): Spatial and Temporal Statistical Analysis of Water Quality Patterns in a Small Temperate Supply Reservoir. *Proceedings of the World Environmental and Water Resources Congress 2012: Crossing Boundaries*. Reston, VA: American Society of Civil Engineers,1982–1992. Available at: http://cedb.asce.org/cgi/WWWdisplay.cgi?290306. Accessed February 2013.
20. Cortina-Borja M, Smith AD, Combarros O, Lehmann D (2009): The synergy factor: A statistic to measure interactions in complex diseases. *BMC Research Notes* 2:105. Available at: http://www.biomedcentral.com/1756-0500/2/105. Accessed December 2012.
21. International Parkinson Disease Genomics C, Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin UM, *et al*. (2011): Imputation of sequence variants for identification of genetic risks for Parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet* 377:641–649.
22. Slatkin M (2008): Exchangeable models of complex inherited diseases. *Genetics* 179:2253–2261.
23. International HapMap C (2003): The International HapMap Project. *Nature* 426:789–796.
24. Ridge PG, Maxwell TJ, Corcoran CD, Norton MC, Tschanz JT, O'Brien E, *et al*. (2012): Mitochondrial genomic analysis of late onset Alzheimer's disease reveals protective haplogroups H6A1A/H6A1B: The Cache County Study on Memory in Aging. *PLoS One* 7:e45134.
25. O'Brien E, Rogers AR, Beesley J, Jorde LB (1994): Genetic structure of the Utah Mormons: Comparison of results based on RFLPs, blood groups, migration matrices, isonymy, and pedigrees. *Hum Biol* 66: 743–759.
26. Jorde LB, Morgan K (1987): Genetic structure of the Utah Mormons: Isonymy analysis. *Am J Phys Anthropol* 72:403–412.
27. Jorde LB (1982): The genetic structure of the Utah Mormons: Migration analysis. *Hum Biol* 54:583–597.
28. Cordell HJ (2002): Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11:2463–2468.
29. Moore JH (2003): The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered* 56:73–82.
30. Tang MX, Maestre G, Tsai WY, Liu XH, Feng L, Chung WY, *et al*. (1996): Relative risk of Alzheimer disease and age-at-onset distributions, based on APOE genotypes among elderly African Americans, Caucasians, and Hispanics in New York City. *Am J Hum Genet* 58:574–584.
31. Murrell JR, Price B, Lane KA, Baiyewu O, Gureje O, Ogunniyi A, *et al*. (2006): Association of apolipoprotein E genotype and Alzheimer disease in African Americans. *Arch Neurol* 63:431–434.
32. Mayeux R (2003): Apolipoprotein E, Alzheimer disease, and African Americans. *Arch Neurol* 60:161–163.
33. Desai PP, Hendrie HC, Evans RM, Murrell JR, DeKosky ST, Kamboh MI (2003): Genetic variation in apolipoprotein D affects the risk of Alzheimer disease in African-Americans. *Am J Med Genet B Neuropsychiatr Genet* 116B:98–101.
34. Maestre G, Ottman R, Stern Y, Gurland B, Chun M, Tang MX, *et al*. (1995): Apolipoprotein E and Alzheimer's disease: Ethnic variation in genotypic risks. *Ann Neurol* 37:254–259.
35. Healy DG (2006): Case–control studies in the genomic era: A clinician's guide. *Lancet Neurol* 5:701–707.
36. Lee SH, Harold D, Nyholt DR, Consortium AN, International Endogene C, Genetic, *et al*. (2013): Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet* 22:832–841.