Alzheimer's & Dementia 12 (2016) 121-129

Featured Article

Interaction between variants in *CLU* and *MS4A4E* modulates Alzheimer's disease risk

Mark T. W. Ebbert[a], Kevin L. Boehme[a], Mark E. Wadsworth[a], Lyndsay A. Staley[a], for the Alzheimer's Disease Neuroimaging Initiative[1], Alzheimer's Disease Genetics Consortium, Shubhabrata Mukherjee[b], Paul K. Crane[b], Perry G. Ridge[a], John S. K. Kauwe[a,*]

[a]*Department of Biology, Brigham Young University, Provo, UT, USA*
[b]*Department of Medicine, University of Washington, Seattle, WA, USA*

Abstract

Introduction: Ebbert et al. reported gene-gene interactions between rs11136000-rs670139 (*CLU-MS4A4E*) and rs3865444-rs670139 (*CD33-MS4A4E*). We evaluate these interactions in the largest data set for an epistasis study.

Methods: We tested interactions using 3837 cases and 4145 controls from Alzheimer's Disease Genetics Consortium using meta-analyses and permutation analyses. We repeated meta-analyses stratified by apolipoprotein E (*APOE*) ε4 status, estimated combined odds ratio (OR) and population attributable fraction (cPAF), and explored causal variants.

Results: Results support the *CLU-MS4A4E* interaction and a dominant effect. An association between *CLU-MS4A4E* and *APOE* ε4 negative status exists. The estimated synergy factor, OR, and cPAF for rs11136000-rs670139 are 2.23, 2.45, and 8.0, respectively. We identified potential causal variants.

Discussion: We replicated the *CLU-MS4A4E* interaction in a large case-control series and observed *APOE* ε4 and possible dominant effect. The *CLU-MS4A4E* OR is higher than any Alzheimer's disease locus except *APOE* ε4, *APP*, and *TREM2*. We estimated an 8% decrease in Alzheimer's disease incidence without *CLU-MS4A4E* risk alleles and identified potential causal variants.

Keywords: Alzheimer's disease; Epistasis; MS4A4E; CLU; CD33; Meta-analysis; ADGC; ADNI

1. Introduction

Alzheimer's disease (AD) is a complex neurodegenerative disease and is the third leading cause of death in the United States [1]. AD is characterized by the accumulation of amyloid plaques and neurofibrillary tangles in the brain. Many genetic loci exist that modify AD risk, but collectively, they explain only a fraction of AD's heritability [2] and are not diagnostically useful [3,4]. Rare variants with large effects and epistatic interactions may account for much of the unexplained AD heritability, but are largely unknown due to limitations in traditional genome-wide association studies. Although rare variants 1and epistatic effects on AD are poorly understood, recent studies suggest that gene-gene interactions play a critical role in AD etiology and progression [3,5–7].

A previous study [3] reported evidence of two gene-gene interactions that increase AD risk. Specifically,

The authors report no conflicts of interest, financial or otherwise.

[1]Data used in preparation of this article were obtained in part from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

*Corresponding author. Tel.: +1-801-422-2993; Fax: +801-422-0900.
E-mail address: kauwe@byu.edu

http://dx.doi.org/10.1016/j.jalz.2015.08.163
1552-5260/

Ebbert et al. reported interactions between rs11136000 C/C (*CLU*; minor allele = T, MAF = 0.38) and rs670139 G/G (*MS4A4E*; minor allele = T, MAF = 0.38) genotypes (synergy factor [SF] = 3.81; P = .016) and the rs3865444 C/C (*CD33*; minor allele = A, MAF = 0.21) and rs670139 G/G (*MS4A4E*) genotypes (SF = 5.31; P = .003). All three variants have been implicated in numerous AD GWAS studies [8–13] and are on the "AlzGene Top Results" list [14], which summarizes the most established genes associated with AD.

MS4A4E and *CLU* were recently replicated in a large meta-analysis of 74,046 individuals, but *CD33* did not replicate [15]. Despite *CD33* failing to replicate, several studies demonstrated that *CD33* is involved in AD-related pathways and pathology, giving convincing evidence that *CD33* is somehow involved in AD. Three specific studies demonstrated that *CD33* alters monocyte function, amyloid uptake, and that *CD33* expression is associated with clinical dementia ratings [16–18]. rs3865444 is located in the 5′ untranslated regions (UTR) of *CD33*.

The association between *CLU* and AD status has been strongly established by both genetic and biological data. Recent studies demonstrated that rs11136000—an intronic single nucleotide polymorphism within *CLU*—is associated with AD-related pathology in healthy individuals including neural inefficiency [19] and decreased white matter integrity [20].

MS4A4E is a member of the membrane-spanning 4-domains subfamily A, but little else is known about the gene. However, rs670139—located in the *MS4A4E* 3′UTR according to gene model XM_011545416.1—is consistently associated with AD [15,18,21].

In this study, we attempted to replicate these gene-gene interactions using the largest data set used in an epistasis study, to date [22]. We performed an independent meta-analysis of data sets from the Alzheimer's Disease Genetics Consortium (ADGC) using 3837 cases and 4145 controls, followed by a combined meta-analysis that included the original Cache County results [3] with an additional 326 cases and 2093 controls. We also tested for dosage or dominant effects and an apolipoprotein E (*APOE*) ε4 effect. Finally, we explored possible causal variants using whole-genome sequence data from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

2. Methods

2.1. Data description

We used SNP data from the ADGC, which consists of 32 studies collected over two phases and includes 16,000 cases and 17,000 controls. All subjects are self-reported as being of European-American ancestry. More information about this data set can be found in the study by Naj et al. [8] and the ADGC data preparation description [23].

Genotype data from 2419 individuals from the Cache County Study on Memory Health and Aging were also used in this study. The full cohort of 5092 individuals represented approximately 90% of the Cache County population aged ≥65 years when the study began in 1994 [24]. The Cache County data consist exclusively of individuals of European-American ancestry. Exactly 2673 individuals were excluded from the original Cache County analysis because of incomplete genotype or clinical data [3]. Additional information on this data set can be found in previous reports [3,24].

Whole-genome data from 747 individuals (223 controls, 195 cases, and 329 mild cognitive impairment [MCI]) were used in this article and were obtained from the ADNI database (adni.loni.usc.edu). ADNI is a large collaboration from several academic and private institutions, and subjects have been recruited from over 50 sites across the United States and Canada. Currently, over 1500 adults (ages 55–90) participate, consisting of cognitively normal older individuals, people with early or late MCI, and people with early stage AD. For up-to-date information, see www.adni-info.org.

2.2. SNP data preparation and statistical analysis

As gene-gene interactions are challenging to identify and replicate, we used the highest quality data possible. For each ADGC data set, we filtered SNPs imputed with low information (info <0.5) and converted the IMPUTE2/SNPTEST format files to PLINK format, using PLINK v1.90b2i [25,26]. We used the default PLINK uncertainty cutoff of 0.1, meaning any imputed call with uncertainty greater than 0.1 was treated as missing. We included SNPs with a missing genotype rate less than 0.05 and individuals with a missing rate less than 0.01. We then extracted the SNPs of interest: rs3865444 (CD33), rs670139 (*MS4A4E*), and rs11136000 (*CLU*) and tested Hardy-Weinberg equilibrium [27,28]. Using R version 3.1.1 [29], we excluded samples without complete data for all covariates including age, gender, case-control status, *APOE* ε4 dose, and the two SNPs being tested in the corresponding interaction. Entire data sets missing in the respective SNPs or covariates after data cleaning were excluded from further analysis. The requirement of complete data for both SNPs and all covariates is necessary for this analysis. Unfortunately, this requirement led to the exclusion of 23 and 24 entire data sets for the *CD33-MS4A4E* and *CLU-MS4A4E* interactions, respectively. We also excluded the ADC1 data set because it contained only one AD case, likely making it biased.

After data preparation, we tested the individual interactions in each data set using logistic regression. We defined the R models as "case_control ~ rs3865444 + rs670139 + rs3865444:rs670139 + apoe4dose + age + sex" and "case_control ~ rs11136000 + rs670139 + rs11136000:rs670139 + apoe4dose + age + sex" for the *CD33-MS4A4E* and *CLU-MS4A4E* interactions, respectively.

Case-control status, SNPs, and sex were coded as factors, age was numeric, and apoe4 dose was an ordered factor from 0 to 2.

Using results from each study, we performed a meta-analysis to test replication across the ADGC data sets using METAL (version 2011-03-25) [30] and performed a second meta-analysis including the original Cache County results to provide SF and odds ratio (OR) estimates from the largest number of samples possible. We tested the originally reported interactions and heterozygous interactions (rs11136000 C/C—rs670139 G/T and rs3865444 C/C—rs670139 G/T) to test for potential dosage or dominant effects based on suggestive evidence found in the original Cache County study (Supplementary Table 1). We assessed whether there is a dosage or dominant effect based both on whether the heterozygous interaction is significant and a *t* test comparing two means. Specifically, we tested for a significant difference between the homozygous and heterozygous effect sizes. A significant difference would suggest a dosage effect, whereas an insignificant difference would suggest the effect might be dominant.

After the meta-analyses, we performed a permutation analysis with 10,000 permutations for interactions that replicated independently. For each ADGC data set, we randomly permuted case-control status across all individuals, tested the interaction, and reran the meta-analysis. We stored the *P* values from each of the 10,000 meta-analyses and calculated the empirical *P* value by finding the original *P* value's rank in the distribution of *P* values divided by the number of permutations. We also calculated the combined population attributable fraction (cPAF) as previously described [3,8].

Results are represented using both ORs and SFs [6,31] and their associated 95% confidence intervals and *P* values. SFs represent the ratio between the "observed and expected ORs" for the two interacting SNPs (Equation 1). The "expected OR" for the interaction assumes there is no synergy between the SNPs (i.e., the SNPs are independent) and equals the product of the individual ORs (the denominator of Equation 1) [6,31]. Essentially, the SF measures how strongly the "observed and expected OR" relationship deviates from linearity, as the SF deviates from 1. A SF = 1 suggests no synergy; rather, there is no evidence of statistical epistasis.

Because SFs <1 can be challenging to interpret, we present interaction SFs in the direction >1. Consequently, we performed all interaction analyses using each gene's homozygous minor allele as the reference group, which is opposite the direction standardly used in genome-wide association studies. This also has the added advantage that the interaction's OR is presented in the risk direction for easy comparison with top AD risk loci. To calculate the interaction's OR, we used each SNPs individual OR as previously reported in a larger data set [21], but we had to invert the individual ORs to be the same direction as our analyses. We then calculated the interaction's "observed OR" (Equation 1) using the inverted ORs. We also estimated each SF's 95% confidence interval using rmeta [32].

Based on results from the interaction replication, we performed a SF analysis using the Cortina-Borja [31] SF calculator to test for an *APOE* ε4 effect for the *CLU-MS4A4E* interaction. Specifically, we stratified the combined ADGC and Cache County data by *APOE* ε4 status and tested for an association between the interaction and case-control status within each stratum. Alleles rs11136000 C and rs670139 G were used as the exposed groups.

2.3. Exploring causal variants

As a follow-up analysis, we explored causal variants for replicated interactions using 747 (223 controls, 195 cases, and 329 MCI) ADNI whole genomes that were sequenced, aligned to hg19, and variants identified by Illumina using their internal analysis procedure. We used linkage disequilibrium, RegulomeDB (accessed November 2014) [33], and functional annotations from wANNOVAR [34] to isolate SNPs of interest. We first extracted all SNPs within approximately 50 kb of each SNP of interest, calculated linkage disequilibrium using Haploview [35], and retained all SNPs with a D' $\geq$0.99. Using RegulomeDB and wANNOVAR, we annotated each remaining SNP for: (1) known regulation and functional effects; (2) minor allele frequencies from the 1000 Genomes Project [36], 6500 Exomes Project [37], and the ADNI data set; and (3) corresponding MutationTaster predictions [38]. We retained all nonsynonymous SNPs, SNPs located in UTRs, and SNPs with a RegulomeDB score <4. For each retained SNP, we tested individual associations with case-control status in the 223 controls and 195 cases using the VarStats tool in the VariantToolChest [39] and subsequently tested their interaction with all SNPs in the other interacting gene using logistic regressions in R.

3. Results

3.1. Sample and data set demographics

Sample demographics and minor allele frequencies for rs11136000, rs670139, and rs3865444 are presented for each data set (Table 1). Eight of the 32 data sets with 3837 cases and 4145 controls passed quality controls for the *CD33-MS4A4E* interaction, whereas 7 data sets with 3140 cases and 2713 controls passed for *CLU-MS4A4E*. The remaining data sets were either missing required SNP(s), missing a covariate, or consisted of only controls and could not be included in the analysis. All SNPs passed Hardy-Weinberg equilibrium in all remaining data sets for both cases and controls.

3.2. Homozygous and heterozygous interaction meta-analysis results

The heterozygous interaction between the rs11136000 C/C (*CLU*) and rs670139 G/T (*MS4A4E*) genotypes did not replicate in the independent analysis, although it is

Table 1
Sample demographics by data set

Study	n	Cases (%)	Females (%)	Age	*APOE* ε4+ (%)	rs670139 MAF (T)	rs3865444 MAF (A)	rs11136000 MAF (T)
ACT1	1858	487 (26.2)	1068 (57.5)	82.28	526 (28.3)	0.41	0.33	NA
ADC2	681	566 (83.1)	365 (53.6)	79.38	394 (57.9)	0.42	0.33	0.39
ADNI	371	230 (62.0)	157 (42.3)	77.82	201 (54.2)	0.45	0.31	0.37
LOAD	2965	1515 (51.1)	1882 (63.5)	78.22	1667 (56.2)	0.43	0.31	0.38
TARC1	388	244 (62.9)	244 (62.9)	78.96	189 (48.7)	0.43	0.32	0.41
UMVUMSSM_A	1058	450 (42.5)	676 (63.9)	75.48	451 (42.6)	0.43	0.31	0.38
UMVUMSSM_B	390	135 (34.6)	236 (60.5)	73.99	118 (30.3)	0.41	0.33	0.38
UMVUMSSM_C	271	210 (77.5)	160 (59.0)	74.77	167 (61.6)	0.42	0.29	NA

Abbreviation: NA, not applicable.

NOTE. For each dataset, the following information is provided: percent cases, females, age, *APOE* ε4 positive percentage, and minor allele frequencies for rs670139, rs3865444, and rs11136000.

suggestive (SF = 1.58, P = .07, Fig. 1A; Supplementary Table 1). Although the heterozygous interaction did not replicate independently in ADGC, the combined meta-analysis including Cache County is significant (SF = 1.90, P = .01, Fig. 1A; Supplementary Table 1). The originally reported homozygous *CLU-MS4A4E* interaction between the rs11136000 C/C (*CLU*) and rs670139 G/G (*MS4A4E*) genotypes replicates in the independent meta-analysis (SF = 1.79, P = .008, Fig. 1B; Supplementary Table 1). The combined meta-analysis is also significant (SF = 2.23, P = .0004, Fig. 1B; Supplementary Table 1). The individual SNP ORs, as previously reported for rs11136000 and rs670139 [21], are 0.83 and 1.09, respectively. The inverted individual SNP ORs for rs11136000 and rs670139 are 1.20 and 0.92, respectively. The expected OR for the interaction is 1.20 × 0.92 = 1.10, thus the observed OR is 2.23 × 1.10 = 2.45. Empirical P values obtained from permutations support the main interaction (ADGC: P = .035 with Cache: P = .002) and the cPAF for *CLU-MS4A4E* is 8.0. Comparing means to determine whether there is a dosage or dominant effect between the heterozygous and homozygous interactions was not significant (P = .22).

We found an association between the *CLU-MS4A4E* interaction and case-control status in *APOEε4* negative subjects in the combined ADGC and Cache County data (SF = 2.08, P = .004, Fig. 2A; Supplementary Table 2) that did not exist with *APOE* ε4 positive subjects (SF = 1.19, P = .26, Fig. 2B; Supplementary Table 2). The *CD33-MS4A4E* interaction failed to replicate in either the independent or combined meta-analyses (Fig. 3A and B, Supplementary Table 1).

3.3. Exploring causal variants

We explored causal variants in the *CLU* and *MS4A4E* regions using the ADNI whole-genome data. There were 36 and 32 SNPs that fit the inclusion criteria previously described for SNPs near rs11136000 and rs670139, respectively (Supplementary Tables 3 and 4). Most of the SNPs

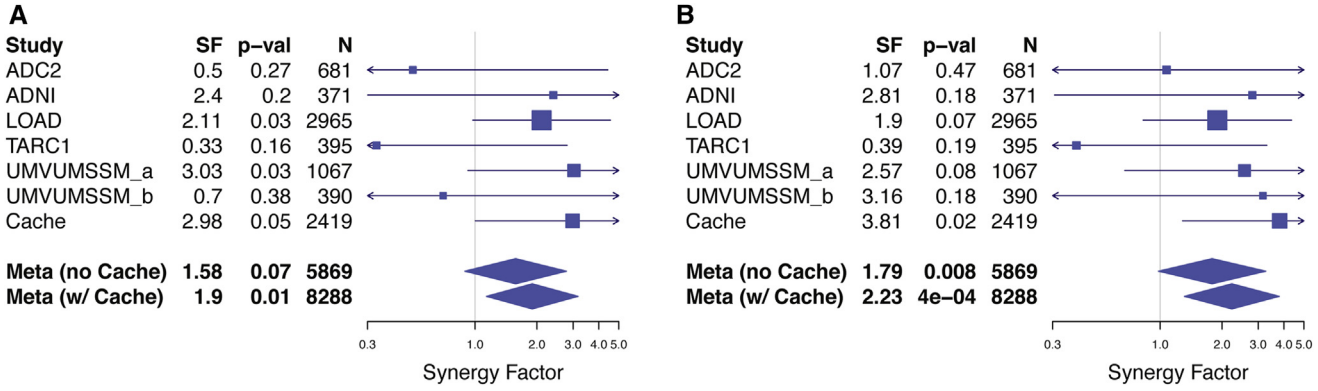

Fig. 1. Forest plot showing heterozygous (panel [A]) and homozygous (panel [B]) CLU-MS4A4E interaction replication with potential dominant effect. We tested the original homozygous interaction between the rs11136000 C/C (CLU; minor allele = T, MAF = 0.38) and rs670139 G/G (MS4A4E; minor allele = T, MAF = 0.38) genotypes, which replicated in ADGC independently (synergy factor = 1.79, P = .008, panel [B]). We also report the combined meta-analysis including the original Cache County results to present a synergy factor estimate from the largest number of samples possible, which is also significant (synergy factor = 2.23, P = 4e-04, panel [B]). We also tested for a dosage or dominant effect based on suggestive evidence in the original Cache County results (panel [A]) by testing the heterozygous interaction between the rs11136000 C/C (CLU) and rs670139 G/T (MS4A4E) genotypes, which did not replicate independently but is suggestive (synergy factor = 1.58, P = .07, panel [A]). The combined analysis, including the original Cache County results, is significant (synergy factor = 1.90, P = .01, panel [A]). Abbreviation: ADGC, Alzheimer's Disease Genetics Consortium.

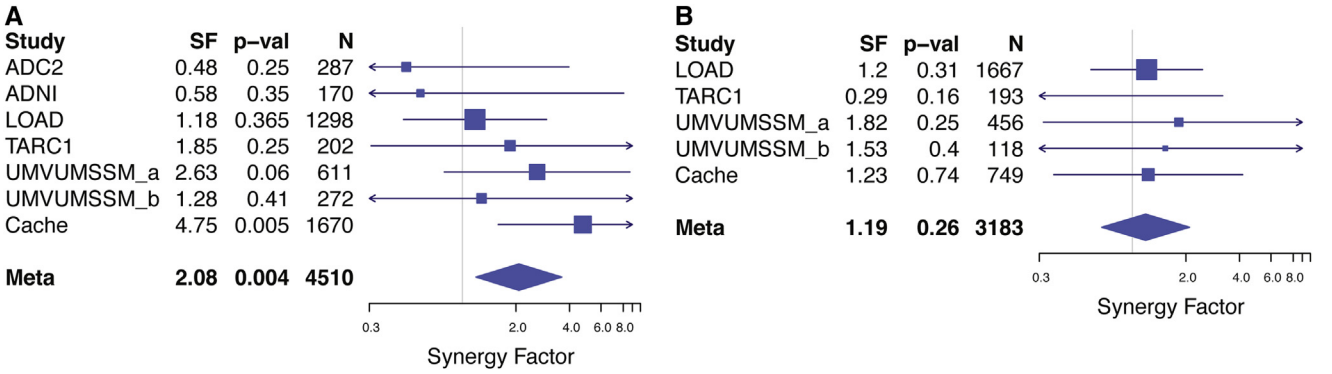

Fig. 2. Forest plot showing *APOE* ε4 negative association with Alzheimer's disease case-control status. We tested for an *APOE* ε4 association with the CLU-MS4A4E interaction using the Cortina-Borja Synergy Factor Calculator [31]. Specifically, we stratified the combined ADGC and Cache County data by *APOE* ε4 status and tested for an association between the interaction and case-control status within each stratum. Alleles rs11136000 C and rs670139 G were used as the exposed groups. We found an association in the *APOE* ε4 negative stratum (synergy factor = 2.08, $P = .004$, panel [A]) that did not exist in the *APOE* ε4 positive stratum (synergy factor = 1.19, $P = .26$, panel [B]), suggesting an *APOE* ε4 effect exists for this interaction. Abbreviation: ADGC, Alzheimer's Disease Genetics Consortium.

are rare (MAF <0.01) according to the 1000 genomes, 6500 exomes, and ADNI data sets. None of the SNPs were significantly associated with case-control status individually or in the pairwise interactions. We identified two SNPs in *MS4A4E* (rs2081547 and rs11230180) that have a RegulomeDB score of "1f" and have been shown to modify *MS4A4A* expression [40], the gene upstream from *MS4A4E*. A score of "1f" means they are known to modify expression and are known DNase and transcription factor binding sites.

4. Discussion

In this study, we attempted to replicate two gene-gene interactions and their association with AD case-control status in the largest data set used in an epistasis study, to date. The *CD33-MS4A4E* interaction failed to replicate and may have resulted from overfitting in the Cache County data as previously described by Ebbert et al. [3] Overfitting happens when a model identifies random data patterns as significant when they are not truly relevant to the question at hand. Although there is substantial evidence that CD33 function is related to AD pathways and pathology [16–18], our data do not support an interaction with *MS4A4E* that impacts AD risk.

We replicated the *CLU-MS4A4E* interaction, demonstrated an association in *APOE ε4* negative subjects, and reported evidence of a possible dominant effect for *MS4A4E*. The homozygous interaction between rs11136000 (*CLU*) and rs670139 (*MS4A4E*) replicates independently in the ADGC data sets, supporting its validity. To provide SF and OR estimates from the largest number of samples possible, we report the combined meta-analysis SF and OR including the Cache County data. Given the broad sampling and large

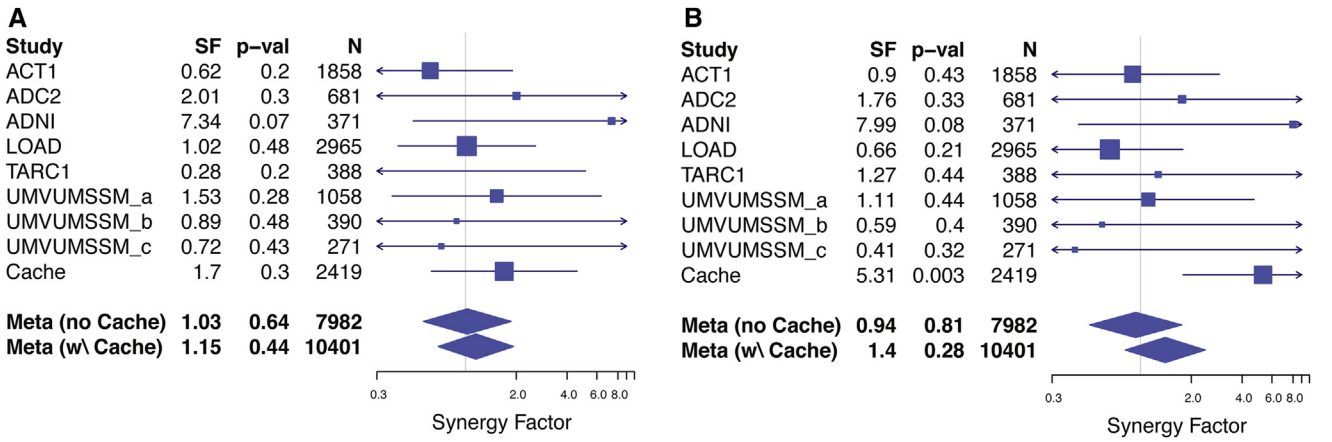

Fig. 3. Forest plot showing failed replication for heterozygous (panel [A]) and homozygous (panel [B]) CD33-MS4A4E interaction. We tested the original homozygous interaction between the rs3865444 C/C (CD33; minor allele = A, MAF = 0.21) and rs670139 G/G (MS4A4E; minor allele = T, MAF = 0.38), which did not replicate in ADGC independently ($P = .81$, panel [B]) and was not significant in the combined meta-analysis ($P = .28$, panel [B]). We also tested the heterozygous interaction, which also was not significant (without Cache: $P = .64$, panel [A]; with Cache: $P = .44$, panel [A]). Abbreviation: ADGC, Alzheimer's Disease Genetics Consortium.

sample size used for this analysis, our results are likely to be generalizable to other populations of European ancestry. Further investigating this interaction in other ethnic groups is warranted.

Comparing the *CLU-MS4A4E* OR of 2.45 to top AD risk alleles according to AlzGene.org [14] along with *APP*, *PLD3*, and *TREM2* adds greater perspective. Momentarily ignoring *APOE* ε4, *APP*, *PLD3*, and *TREM2*, the highest individual OR is from *APOE* ε2 (OR = 1.61) [3,14] when inverting to its respective risk allele, followed by *ABCA7* (OR = 1.23) [3,14], both of which are dramatically lower than 2.45. Of known AD risk loci, only *APOE* ε4 (OR = 3.68), *APP* (OR = 5.29), and *TREM2* (OR = 5.05) have ORs greater than 2.45 [3,14,41,42]. The *CLU-MS4A4E* OR is even greater than the *PLD3* Val232Met mutation (OR = 2.10) [43]. These results suggest the *CLU-MS4A4E* interaction may play an important role in AD etiology.

A distinction must be made regarding statistical and biological epistasis, however [22]. Although there is evidence that CLU, like CD33, interacts indirectly with MS4A2 [3], little is known about MS4A4E itself and we do not know whether it biologically interacts with CLU. MS4A2 indirectly modifies BCL2L1 activation or expression [3], which physically interacts with CLU. Research suggests CLU prevents amyloid fibrils and other protein aggregation events [44], whereas MS4A4E may facilitate aggregation as a membrane-spanning protein. Membrane-spanning proteins play diverse roles in cell activity including transport and signaling. Experiments will be required to determine whether there is a biological epistasis between CLU and MS4A4E, and whether the interaction affects amyloid fibril formation. Our results indicate further investigative efforts in gene-gene interactions (and protein-protein interactions) may be important to resolve AD etiology.

Comparing means for the effect estimates to assess whether there is a dosage or dominant effect between the *CLU-MS4A4E* heterozygous and homozygous interactions was not significant, suggesting there may be a dominant effect for the rs670139 G allele. A dominant effect has important epidemiologic and heritability implications. Because the *CLU-MS4A4E* interaction increases risk, heterozygous individuals may be at equal risk compared with homozygous individuals.

We found an association between the *CLU-MS4A4E* interaction and case-control status in *APOE* ε4 negative subjects in the combined ADGC and Cache County data that did not exist with *APOE* ε4 positive subjects. This potential three-way interaction may provide valuable insight into AD risk and protective factors. A recent article by Jun et al. [45] found *CLU* has a stronger association in *APOE* ε4 positive individuals, whereas the region surrounding *MS4A4E* has a stronger association in *APOE* ε4 negative individuals. Further statistical and biological studies will be necessary to clarify these potential associations. Because all analyses in this study used each gene's homozygous minor allele as the reference group, the interaction between *CLU-MS4A4E* major alleles is framed as a risk factor, meaning the interaction between the minor alleles is protective. Because the tested heterozygous interaction also increases risk, the protective association may only apply to the interaction between the homozygous minor alleles.

We report several rare potential causal variants linked to rs11136000 or rs670139 with a D' $\geq$0.99 in the ADNI whole-genome data. No individual variants were significantly associated with AD risk in the ADNI data, but the analysis was likely underpowered with only 240 controls and 202 cases. Two particularly interesting variants, rs11230180 and rs2081547, are known to affect *MS4A4A* expression. We believe further analysis of these variants is necessary to better understand their involvement in AD. Exploring the effects of rs670139, itself, may also be important. Little is known about *MS4A4E*, including the gene's chromosomal structure. According to gene model XM_011545416.1, rs670139 is in the *MS4A4E* 3′UTR, but other gene models differ. 3′UTR variants can affect transcription and translation.

The cPAF for *CLU-MS4A4E* is 8.0, suggesting there would be an approximate 8% decrease in AD incidence across the population if both major alleles were eliminated. In reality, this estimate is for the causal variants that rs670139 and rs11136000 may be tagging, but the overall effect is nontrivial. Identifying a targeted treatment in the associated pathways could have a significant impact.

A major gap in AD literature to date is the lack of known causal variants. Several SNPs have repeatedly turned up in genome-wide association studies, but the tagSNPs themselves are unlikely to play a direct role in AD etiology. What is more likely is that the tagSNPs are in close linkage disequilibrium with one or more causal variants. We hypothesize two possible explanations as follows: (1) the SNPs are linked to multiple rare variants that drive AD development and progression or (2) there is another common variant in the region with functional effects that remain unknown. In either case, given the biological complexity of AD and results presented in this study, we believe epistasis plays a critical role in AD etiology. As such, the community must continue to identify and vet these and other interactions that are supported in the literature.

$$SF = \frac{OR_{12}}{OR_1 \times OR_2}$$

Equation 1. The synergy factor describes the relationship between the expected odds ratio (denominator) and the observed odds ratio (numerator) for interacting variants. The expected odds ratio (denominator) assumes that both variants are independent (i.e., there is no synergistic, or nonlinear effect on the phenotype), whereas the observed odds ratio (numerator) is the

actual effect. A synergy factor that deviates from 1 indicates a statistical interaction between the variants that affects the phenotype.

Acknowledgments

The authors thank the participants and staff of the many centers that were involved in data collection for their important contributions to this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article. The authors thank Drs. D. Stephen Snyder and Marilyn Miller from NIA who are ex-officio ADGC members.

This work was supported by grants from NIH (R01AG11380, R01AG21136, R01AG31272, R01AG042611, and R01 AG 042437), the Alzheimer's Association (MNIRG-11-205368), the Utah Science, Technology, and Research Initiative, the Utah State University Agricultural Experiment Station, the McMillan Family Trust, and the Brigham Young University Gerontology Program.

The National Institutes of Health, National Institute on Aging (NIH-NIA) also supported this work through the following grants: ADGC, U01 AG032984, RC2 AG036528; NACC, U01 AG016976; NCRAD, U24 AG021886; NIA LOAD, U24 AG026395, U24 AG026390; NIAGADS U24 AG041689; Banner Sun Health Research Institute P30 AG019610; Boston University, P30 AG013846, U01 AG10483, R01 CA129769, R01 MH080295, R01 AG017173, R01 AG025259, and R01AG33193; Columbia University, P50 AG008702, and R37 AG015473; Duke University, P30 AG028377 and AG05128; Emory University, AG025688; Group Health Research Institute, UO1 AG006781, UO1 HG004610, UO1 HG006375; Indiana University, P30 AG10133; Johns Hopkins University, P50 AG005146 and R01 AG020688; Massachusetts General Hospital, P50 AG005134; Mayo Clinic, P50 AG016574; Mount Sinai School of Medicine, P50 AG005138 and P01 AG002219; New York University, P30 AG08051, MO1RR00096, UL1 RR029893, 5R01AG012101, 5R01AG022374, 5R01AG013616, 1RC2AG036502, and 1R01AG035137; Northwestern University, P30 AG013854; Oregon Health & Science University, P30 AG008017 and R01 AG026916; Rush University, P30 AG010161, R01 AG019085, R01 AG15819, R01 AG17917, and R01 AG30146; TGen, R01 NS059873; University of Alabama at Birmingham, P50 AG016582 and UL1RR02777; University of Arizona, R01 AG031581; University of California, Davis, P30 AG010129; University of California, Irvine, P50 AG016573, P50, P50 AG016575, P50 AG016576, P50 AG016577; University of California, Los Angeles, P50 AG016570; University of California, San Diego, P50 AG005131; University of California, San Francisco, P50 AG023501 and P01 AG019724; University of Kentucky, P30 AG028383 and AG05144; University of Michigan, P50 AG008671; University of Pennsylvania, P30 AG010124; University of Pittsburgh, P50 AG005133, AG030653, AG041718, AG07562, and AG02365; University of Southern California, P50 AG005142; University of Texas Southwestern, P30 AG012300; University of Miami, R01 AG027944, AG010491, AG027944, AG021547, and AG019757; University of Washington, P50 AG005136; University of Wisconsin, P50 AG033514; Vanderbilt University, R01 AG019085; and Washington University, P50 AG005681 and P01 AG03991. The Kathleen Price Bryan Brain Bank at Duke University Medical Center is funded by National Institute of Neurological Disorders and Stroke grant # NS39764, National Institute of Mental Health MH60451, and by GlaxoSmithKline. Genotyping of the TGEN2 cohort was supported by Kronos Science. The TGen series was also funded by National Institute on Aging grant AG041232 the Banner Alzheimer's Foundation, the Johnnie B. Byrd Sr. Alzheimer's Institute, the Medical Research Council, and the state of Arizona and also includes samples from the following sites: Newcastle Brain Tissue Resource (funding via the Medical Research Council, local NHS trusts and Newcastle University), MRC London Brain Bank for Neurodegenerative Diseases (funding via the Medical Research Council), South West Dementia Brain Bank (funding via numerous sources including the Higher Education Funding Council for England [HEFCE], Alzheimer's Research Trust [ART], BRACE as well as North Bristol NHS Trust Research and Innovation Department and DeNDRoN), The Netherlands Brain Bank (funding via numerous sources including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, Internationale Stiching Alzheimer Onderzoek), Institut de Neuropatologia, Servei Anatomia Patologica, Universitat de Barcelona. ADNI Funding for ADNI is through the Northern California Institute for Research and Education by grants from Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Company, Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., Alzheimer's Association, Alzheimer's Drug Discovery Foundation, the Dana Foundation, and by the National Institute of Biomedical Imaging and Bioengineering and National Institute on Aging grants U01 AG024904, RC2 AG036535, and K01 AG030514. P.G.R. is supported by Wellcome Trust, Howard Hughes Medical Institute, and the Canadian Institutes of Health Research.

Support was also from the Alzheimer's Association (LAF, IIRG-08-89720; MP-V, IIRG-05-14147) and the US Department of Veterans Affairs Administration, Office of Research and Development, Biomedical Laboratory Research Program.

Whole-genome data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by

the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; Euroimmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern Rev December 5, 2013 California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuroimaging at the University of Southern California.

The authors further acknowledge Dr. Donald Lehmann and Dr. Mario Cortina-Borja for their kind help with the synergy factor analyses.

This study was made possible by the Texas Alzheimer's Research and Care Consortium (TARCC) funded by the state of Texas through the Texas Council on Alzheimer's Disease and Related Disorders. Investigators from TARCC include Baylor College of Medicine: Rachelle Doody, Mimi M. Dang, Valory Pavlik, Wen Chan, Paul Massman, Eveleen Darby, Monica Rodriguear, and Aisha Khaleeq; Texas Tech University Health Sciences Center: Chuang-Kuo Wu, Matthew Lambert, Victoria Perez, and Michelle Hernandez; University of North Texas Health Science Center: Thomas Fairchild, Janice Knebl, Sid E. O'Bryant, James R. Hall, Leigh Johnson, Robert C. Barber, Douglas Mains, Lisa Alvarez, and Rosemary McCallum; University of Texas Southwestern Medical Center: Perrie Adams, Munro Cullum, Roger Rosenberg, Benjamin Williams, Mary Quiceno, Joan Reisch, Ryan Huebinger, Natalie Martinez, and Janet Smith; University of Texas Health Science Center–San Antonio: Donald Royall, Raymond Palmer, and Marsha Polk; Texas A&M University Health Science Center: Farida Sohrabji, Steve Balsis, and Rajesh Miranda; Essentia Institute of Rural Health: Stephen C. Waring; University of North Carolina: Kirk C. Wilhelmsen, Jeffrey L. Tilson, and Scott Chasse.

Supplementary data

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.jalz.2015.08.163.

RESEARCH IN CONTEXT

1. Systematic review: We replicated the *CLU-MS4A4E* interaction (odds ratio [OR] = 2.45, synergy factor = 2.23, P = .0004) in the largest data set used in an epistasis study, to date demonstrating an association in *APOE* ε4 negative subjects, and reported evidence of a possible dominant effect for *MS4A4E* on Alzheimer's disease risk.
2. Interpretation: This association represents a rare result in the study of epistasis in Alzheimer's disease: a strong effect, replicated in multiple independent data sets. Comparing the *CLU-MS4A4E* OR of 2.45 to ORs for each top Alzheimer's disease risk allele along with *APP*, *PLD3*, and *TREM2* adds greater perspective to the interaction's effect. Of well-established Alzheimer's disease risk loci, only *APOE* ε4 (OR = 3.68), *APP* (OR = 5.29), and *TREM2* (OR = 5.05) have ORs greater than 2.45. The OR for the *CLU-MS4A4E* interaction is even greater than the Val232Met mutation in *PLD3* (OR = 2.10).
3. Future directions: These results suggest the *CLU-MS4A4E* interaction may play an important role in Alzheimer's disease.

References

[1] James BD, Leurgans SE, Hebert LE, Scherr PA, Yaffe K, Bennett DA. Contribution of Alzheimer disease to mortality in the United States. Neurology 2014;82:1045–50.
[2] Ridge PG, Mukherjee S, Crane PK, Kauwe JSK, Alzheimer's Disease Genetics Consortium. Alzheimer's disease: Analyzing the missing heritability. PLoS One 2013;8:e79771.
[3] Ebbert MTW, Ridge PG, Wilson AR, Sharp AR, Bailey M, Norton MC, et al. Population-based Analysis of Alzheimer's disease risk alleles implicates genetic interactions. Biol Psychiatry 2014;75:732–7.
[4] Sleegers K, Bettens K, De Roeck A, Cauwenberghe CV, Cuyvers E, Verheijen J, et al. A 22-single nucleotide polymorphism Alzheimer risk score correlates with family history, onset age, and cerebrospinal fluid Aβ42. Alzheimers Dement 2015;11:1452–60.
[5] Bullock JM, Medway C, Cortina-Borja M, Turton JC, Prince JA, Ibrahim-Verbaas CA, et al. 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 2013;34:1309.e1–7.
[6] Combarros O, Cortina-Borja M, Smith AD, Lehmann DJ. Epistasis in sporadic Alzheimer's disease. Neurobiol Aging 2009;30:1333–49.
[7] Kauwe JSK, Bertelsen S, Mayo K, Cruchaga C, Abraham R, Hollingworth P, et al. Suggestive synergy between genetic variants in TF and HFE as risk factors for Alzheimer's disease. Am J Med Genet Part B Neuropsychiatr Genet 2010;153B:955–9.
[8] Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1

are associated with late-onset Alzheimer's disease. Nat Genet 2011; 43:436–41.
[9] Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 2009;41:1088–93.
[10] Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al., CHARGE Consortium, GERAD1 Consortium, EADI1 Consortium. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 2010;303:1832–40.
[11] Hu X, Pickering E, Liu YC, Hall S, Fournier H, Katz E, et al., the Alzheimer's Disease Neuroimaging Initiative. Meta-Analysis for genome-wide association study identifies multiple variants at the BIN1 locus associated with late-onset Alzheimer's disease. PLoS One 2011; 6:e16616.
[12] Guerreiro RJ, Beck J, Gibbs JR, Santana I, Rossor MN, Schott JM, et al. Genetic variability in CLU and its association with Alzheimer's disease. PLoS One 2010;5:e9510.
[13] Bettens K, Brouwers N, Van Miegroet H, Gil A, Engelborghs S, De Deyn PP, et al. Follow-up study of susceptibility loci for Alzheimer's disease and onset age identified by genome-wide association. J Alzheimers Dis 2010;19:1169–75.
[14] Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. Nat Genet 2007;39:17–23.
[15] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al, European Alzheimer's Disease Initiative (EADI), Genetic and Environmental Risk in Alzheimer's Disease (GERAD), Alzheimer's Disease Genetic Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013;45:1452–8.
[16] Bradshaw EM, Chibnik LB, Keenan BT, Ottoboni L, Raj T, Tang A, et al., The Alzheimer Disease Neuroimaging Initiative. CD33 Alzheimer's disease locus: Altered monocyte function and amyloid biology. Nat Neurosci 2013;16:848–50.
[17] Griciuc A, Serrano-Pozo A, Parrado AR, Lesinski AN, Asselin CN, Mullin K, et al. Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. Neuron 2013;78:631–43.
[18] Karch CM, Jeng AT, Nowotny P, Cady J, Cruchaga C, Goate AM. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. PLoS One 2012;7:e50976.
[19] Lancaster TM, Brindley LM, Tansey KE, Sims RC, Mantripragada K, Owen MJ, et al. Alzheimer's disease risk variant in CLU is associated with neural inefficiency in healthy individuals. Alzheimers Dement 2014;11:1144–52.
[20] Braskie MN, Jahanshad N, Stein JL, Barysheva M, McMahon KL, de Zubicaray GI, et al. Common Alzheimer's disease risk variant within the CLU gene affects white matter microstructure in young adults. J Neurosci 2011;31:6764–70.
[21] Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, et al, Alzheimer's Disease Neuroimaging Initiative, CHARGE consortium, EADI1 consortium. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 2011;43:429–35.
[22] Ebbert MT, Ridge PG, Kauwe JS. Bridging the gap between statistical and biological epistasis in Alzheimer's disease. Biomed Res Int 2015; 2015:870123.
[23] Boehme KL, Mukherjee S, Crane PK, Kauwe JSK. ADGC 1000 Genomes combined data workflow. Available at: kauwelab.byu.edu/Portals/22/adgc_combined_1000G_09192014.pdf. Accessed October 12, 2015.
[24] Breitner JC, Wyse BW, Anthony JC, Welsh-Bohmer KA, Steffens DC, Norton MC, et al. APOE-epsilon4 count predicts age when prevalence of AD increases, then declines: the Cache County Study. Neurology 1999;53:321–31.
[25] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559–75.
[26] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience 2015;4:7.
[27] Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:887–93.
[28] Graffelman J, Moreno V. The mid p-value in exact tests for Hardy-Weinberg equilibrium. Stat Appl Genet Mol Biol 2013;12:433–48.
[29] R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
[30] Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genome wide association scans. Bioinformatics 2010; 26:2190–1.
[31] Cortina-Borja M, Smith AD, Combarros O, Lehmann D. The synergy factor: A statistic to measure interactions in complex diseases. BMC Res Notes 2009;2:105.
[32] Lumley T. rmeta: Meta-analysis R package version 2.16. 2012 Available at: http://CRAN.R-project.org/package=rmeta. Accessed October 12, 2015.
[33] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012;22:1790–7.
[34] Chang X, Wang K. wANNOVAR: Annotating genetic variants for personal genomes via the web. J Med Genet 2012;49:433–6.
[35] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 2005; 21:263–5.
[36] Consortium T 1000 GP. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
[37] NHLBI GO. Exome sequencing project (ESP). Available at: http://evs.gs.washington.edu/EVS/. Accessed October 12, 2015.
[38] Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010;7:575–6.
[39] Ebbert MT, Wadsworth ME, Boehme KL, Hoyt KL, Sharp AR, O'Fallon BD, et al. Variant Tool Chest: An improved tool to analyze and manipulate variant call format (VCF) files. BMC Bioinformatics 2014;15:S12.
[40] Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. PLoS One 2010;5:e10693.
[41] Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 2012;488:96–9.
[42] Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. N Engl J Med 2013;368:117–27.
[43] Sassi C, Bras J, Gibbs JR, Hernandez DG, Lupton MK, Powell J, et al, UK Brain Expression Consortium, Alzheimer's Research UK Consortium. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 2014;505:550–4.
[44] Yerbury JJ, Poon S, Meehan S, Thompson B, Kumita JR, Dobson CM, et al. The extracellular chaperone clusterin influences amyloid formation and toxicity by interacting with prefibrillar structures. FASEB J 2007;21:2312–22.
[45] Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert JC, Chung J, Naj AC, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. Mol Psychiatry 2016;21:108–17.