A coding variant in CRI interacts with APOE-ε4 to influence cognitive decline

Brendan T. Keenan1,2, Joshua M. Shulman1,2,3,4, Lori B. Chibnik1,2,3, Towfique Raj1,2,3, Dong Tran1,2, Mert R. Sabuncu5,6, The Alzheimer’s Disease Neuroimaging Initiative6, April N. Allen7, Jason J. Corneveaux7, John A. Hardy8,9, Matthew J. Huentelman7, Cynthia A. Lemere3,10, Amanda J. Myers11,12, Anne Nicholson-Weller3,13, Eric M. Reiman7,14, Denis A. Evans15,16, David A. Bennett16, Philip L. De Jager1,2,3,*

1 Program in Translational Neuropyschiatric Genomics, Department of Neurology, Brigham and Women’s Hospital, 77 Avenue Louis Pasteur, NRB168, Boston, MA 02115
2 Program in Medical and Population Genetics, Broad Institute, 7 Cambridge Center, Cambridge, MA 02142
3 Harvard Medical School, Boston, MA 02115
4 Clinical Investigator Training Program: Beth Israel Deaconess Medical Center – Harvard/MIT Health Sciences and Technology, in collaboration with Pfizer Inc. and Merck & Co.
5 Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, 149 Thirteenth Street, Suite 2301, Charlestown, MA 02129
6 Computer Science and Artificial Intelligence Laboratory (CSAIL) and Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, 32 Vassar St., Cambridge, MA 02139
7 Neurogenomics Division, Translational Genomics Research Institute and Arizona Alzheimer’s Consortium, 445 N. Fifth St., Phoenix, AZ 85004
8 Laboratory of Neurogenetics, National Institute on Aging, Porter Neuroscience Research Center, Bldg. 35, Room 1A1000, Bethesda, MD 20892
9 Reta Lila Weston Laboratories, Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London, England
10 Center for Neurologic Diseases, Brigham and Women’s Hospital, Harvard Medical School, 77 Avenue Louis Pasteur, 7th Floor, Boston, MA 02115
11 Department of Psychiatry and Behavioral Sciences, University of Miami, Miller School of Medicine, Miami, FL 33136
12 Johnnie B. Byrd Sr. Alzheimer’s Center and Research Institute, 4001 East Fletcher Ave., Tampa, FL 33136
13 Beth Israel Deaconess Medical Center, 330 Brookline Ave., Boston, MA 02115
14 Banner Alzheimer’s Institute and Department of Psychiatry, University of Arizona, 901 East Willetta St., Phoenix, AZ 85006
15 Departments of Internal Medicine and Neurological Science, Rush University Medical Center, 600 S Paulina St., Chicago, IL 60612
16 Rush Alzheimer’s Disease Center, Rush University Medical Center, 600 S Paulina St., Chicago, IL 60612

*Correspondence: Philip L. De Jager, MD, PhD
Program in Translational NeuroPsychiatric Genomics
Department of Neurology
Brigham and Women’s Hospital
77 Avenue Louis Pasteur, NRB 168C
Boston, MA 02115
Phone: 617 525-4529
Fax: 617 525-5333
Email: pdejager@rics.bwh.harvard.edu

© The Author 2012. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
Abstract

Complement Receptor 1 (CR1) is an Alzheimer’s disease (AD) susceptibility locus that also influences AD-related traits such as episodic memory decline and neuritic amyloid plaque deposition. We implemented a functional fine-mapping approach, leveraging intermediate phenotypes to identify functional variant(s) within the CR1 locus. Using 1709 subjects (697 deceased) from the Religious Orders Study and the Rush Memory and Aging Project, we tested 41 SNPs within the linkage disequilibrium block containing the published CR1 AD SNP (rs6656401) for associations with episodic memory decline, and then examined the functional consequences of the top result. We report that a coding variant in the LHR-D region of the CR1 gene, rs4844609 (Ser1610Thr, MAF=0.02), is associated with episodic memory decline and accounts for the known effect of the index SNP rs6656401 (D’=1, r²=0.084) on this trait. Further, we demonstrate that the coding variant’s effect is largely dependent on an interaction with APOE-ε4 and mediated by an increase burden of AD-related neuropathology. Finally, in our data, this coding variant is also associated with AD susceptibility (joint odds ratio=1.4). Taken together, our analyses identify a CR1 coding variant that influences episodic memory decline; it is a variant known to alter the conformation of CR1 and points to LHR-D as the functional domain within the CR1 protein that mediates the effect on memory decline. We thus implicate C1q and MBL, which bind to LHR-D, as likely targets of the variant’s effect and suggest that CR1 may be an important intermediate in the clearance of Aβ42 particles by C1q.
Introduction

Alzheimer’s disease (AD) is among the most common age-related conditions, affecting one out of every eight US adults over age 65 (1) with progressive amnesia, dementia, global cognitive failure, and ultimately death. Recent large genome-wide association studies (GWAS) have both identified and replicated several new AD susceptibility loci outside of Apolipoprotein E (APOE), including complement receptor 1 (CR1), which is the primary focus of this manuscript (2-7).

CR1 is a large, polymorphic, and multifunctional glycoprotein composed primarily of independently folding domains, known as complement control protein repeats (CCPs), each containing 59-75 amino acids (8, 9). The CR1 gene (Figure 1) contains many different polymorphisms; its most common version contains 30 CCPs that form 4 long homologous repeats (LHR-A, LHR-B, LHR-C, and LHR-D), each composed of 7 CCPs (8, 9). CR1 has been postulated to impact AD pathogenesis through its key role in regulating complement activity as a receptor of the complement C3b protein (4). Complement activation leads to a proteolytic cascade, which produces activated C3b and C4b fragments that bind covalently to acceptor molecules and mediate phagocytosis through CR1 (4). Some investigators have suggested that this mechanism may mediate CR1’s clearance of Aβ particles, which can aggregate to form amyloid plaques, an important component of AD neuropathology (4). The involvement of CR1 in the clearance of soluble Aβ particles may also be mediated through its role as an adherence receptor on red blood cells for immobilizing complement tagged particles in the blood and delivering these particles for removal by resident phagocytes of the liver and spleen (10, 11). Our earlier work associating the reported AD susceptibility allele in CR1 with a greater burden of amyloid pathology (12) is consistent with these reports and suggests that the CR1 locus may influence AD-related molecular processes that are also modulated by the well established APOE
susceptibility locus (13, 14).

Our analyses focus on the region surrounding the published single nucleotide polymorphism (SNP) rs6654601 in the CR1 locus (4), for which we have previously described an association with decline in episodic memory and also the burden of neuritic amyloid plaques (12). Since the CR1 locus is implicated in both AD susceptibility and intermediate traits related to AD, we explored the implicated chromosomal segment that contains most of the CR1 gene for additional SNPs with evidence of being associated with these intermediate traits. Identifying genetic variation that influences the function of a locus will inform our understanding of the targeted gene’s biological function, and some of these functional variants may also be involved in disease susceptibility. Based on our previous study (12) and the fact that we have a large sample size for clinical intermediate phenotypes, we selected decline in episodic memory, a key feature of AD (15), as our primary outcome measure with which to identify additional variants of interest within the CR1 locus. We then validated our results and extended our analyses to include other pertinent traits, including measures of AD pathology.

Results

Demographics. Descriptive characteristics for the Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP) subjects are presented in Supporting Information (SI) Table S1; at the time of enrollment, all subjects are non-demented. Genetic information was available for 817 subjects in ROS and 892 subjects from MAP. Of these 1709 subjects, an estimate of episodic memory decline was available for the 1593 (93.2%) with at least two measures of episodic memory. A total of 697 (40.7%) subjects were deceased, and 651 (93.4%) subjects had available neuritic amyloid plaque and neurofibrillary tangle pathology data at the time of our
analyses. The average age at enrollment for ROS was about five years younger than for MAP (76 vs. 81 years), and the average age at death was nearly two years younger for ROS (87 vs. 89 years). Twenty percent (n=340) of participants were clinically diagnosed with AD during the course of study.

Descriptive characteristics for subjects from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and the Chicago Health and Aging Project (CHAP) are listed in SI Table S2. Of the 746 individuals for whom data was available in ADNI, 207 (27.8%) were cognitively normal, 365 (48.9%) had mild cognitive impairment, and 174 (23.3%) had probable AD at baseline. A total of 311 (41.7%) had clinically diagnosed AD at their last recorded follow up. Episodic memory decline was available in 716 (96%) of the samples. The mean (standard deviation (SD)) age was 75.4 (6.9), the mean years of education was 15.6 (3.0) and 306 (41.0%) were female. Our CHAP sample consisted of 624 European American subjects, 569 (91.2%) of whom had both genetic information and at least 2 measures of episodic memory. At the time of our sample, 50 (11.6%) had a diagnosis of AD, the mean (SD) age was 71.9 (5.2), the mean years of education was 14.9 (3.3) and 393 (63.0%) were female. Finally, our case-control sample from the Translational Genetics Research Institute (TGEN) consisted of 1004 pathologically confirmed AD cases and 575 controls with both genetic and clinical information. All subjects were included in a recent GWAS for AD susceptibility (2). The mean (SD) age at death was 82.1 (7.7) for cases and 80.9 (8.7) for controls and 64% of cases and 48% of controls were female.

Decline in Episodic Memory. Our primary analysis examined the associations between episodic memory decline and 41 SNPs within the block of linkage disequilibrium (LD) containing the index CR1 SNP, rs6656401, that is associated with AD susceptibility (2, 4). These SNPs were
the ones that met quality criteria (info score >0.3 and minor allele frequency (MAF) >0.01) at the end of our imputation pipeline and were therefore available in our dataset of imputed allele doses generated using the HapMap 2 reference (16). Figure 2 and SI Figure S1 give a more comprehensive view of the chromosomal segment containing CR1, illustrating the associations for all SNPs within a 500kb radius of rs6656401. As previously reported (12), we note a significant association with the index CR1 SNP rs6656401 (p=0.048), such that, for each additional rs6656401^A allele, there is a decrease of 0.009 standard unit of episodic memory per year (Table 1). The most associated SNP in the CR1 LD block is the non-synonymous coding SNP (nscSNP) rs4844609 (p=0.003), which results in the substitution of a Threonine for a Serine residue at amino acid position 1610 (Figure 2). This result is significant since the 41 SNPs tested in the CR1 LD block actually reflect only 10 independent groups of SNPs in strong LD (r^2>0.5); a p<0.005 is therefore our threshold of significance after Bonferroni correction. This nscSNP (rs4844609) lies in exon 37, which encodes a segment of the LHR-D region of the CR1 gene, and has only been observed in European (MAF of 2%) and East Asian populations (MAF of 1-5%) (http://hapmap.ncbi.nlm.nih.gov). The Rsq quality metric (the squared correlation between imputed and true genotypes) for rs4844609 was 0.7212 in ROS and MAP, compared to an Rsq of 0.9997 for rs6656401. Persons with the rs4844609^A susceptibility allele had a fourfold larger decrease in standardized episodic memory per year when compared to subjects with the original susceptibility allele rs6656401^A (Table 1). A meta-analysis within two additional sample collections (ADNI and CHAP) confirms the association between rs4844609^A and episodic memory decline (p=0.004) (Table 2). The Rsq quality metric for rs4844609 and rs6656401, respectively, in these samples was 0.6868 and 0.9091 in CHAP and 0.1426 and 0.9270 in ADNI. We do not see any significant association for the index CR1 SNP, rs6656401, in the replication
samples. We therefore consider rs4844609 a possible functional variant that influences episodic memory decline within the *CR1* locus.

To determine whether the two variants have independent effects on our trait of interest, we next fit a joint regression model containing individual terms for both rs6656401 and rs4844609, along with the relevant covariates (Table 1). We see that after adjusting for the effect of rs4844609, the observed association between the index *CR1* SNP (rs6656401) and episodic memory decline decreases by more than half and is no longer significant. By contrast, the observed association with our functional variant (rs4844609) was only reduced by 10% and remained significant after adjusting for the rs6656401. These results suggest that the association of episodic memory decline with the *CR1* locus is best captured by the rs4844609 variant.

This observation is illustrated in Figure 3, which presents the mean trajectory of different genotypic subject classes based on rs6656401 and rs4844609 genotypes. As previously reported, the effect of the index SNP rs6656401 is best captured using a dominant model (12), which is used in these illustrations. There are no rs4844609AA homozygotes in our dataset. We see that all rs4844609A risk alleles are found in subjects with at least one copy of the rs6656401A allele, and that these subjects decline significantly faster than those with neither risk allele (p=0.013). We also note that the intercept of the mean cognitive trajectory of subjects with an rs4844609A allele is significantly lower than in the other genotype categories (p=0.010) (Figure 3). This observation suggests that decline in episodic memory related to rs4844609A has probably been ongoing for many years prior to the start of the study. Illustrations of mean episodic decline for each SNP separately are presented in SI Figure S2.

In secondary analyses, we explored the hypothesis that the effect of *CR1* on episodic memory may change over the course of aging; specifically, we assessed whether its effect may be
different in non-AD and AD subjects. Given the stratification of subjects, our statistical power is limited in this secondary analysis. We report that the magnitude of the effect of the rs4844609^A allele on the decline in episodic memory is similar in both groups of subjects (SI Table S3); however, the analyses are not significant given the reduced sample size. So, while this result suggests that CR1 may affect decline throughout the course of aging, a definitive answer to this question will require larger sample sizes.

**Interaction with APOE-ε4.** Confirming and extending previous reports for the CR1 locus (4), we find a significant interaction between our candidate functional variant rs4844609 and the presence or absence of APOE-ε4 on episodic memory decline (p=0.027) (SI Table S4). Specifically, among carriers of APOE-ε4, subjects with the rs4844609^A risk allele have a decrease of 0.084 standard unit of episodic memory per year more than those with no rs4844609^A alleles (p=0.005). We do not observe a significant association for rs4844609^A among non APOE-ε4 carriers, nor do we observe significant interaction between APOE-ε4 and the index CR1 SNP rs6656401^A. As shown in SI Table S4 and illustrated in Figure 4, we observe a more rapid decline in episodic memory for APOE-ε4 carriers with the rs4844609^A risk allele, compared to non APOE-ε4 carriers with no rs4844609^A risk alleles (p<0.0001). We note that the slope of decline in the strata with both APOE-ε4 and the rs4844609^A risk allele is significantly steeper than the other 3 groups (p≤0.005 for all comparisons), although this is based on a small number of subjects with both risk variants.

**Clinically Diagnosed AD.** Given the robust effect on decline in episodic memory, we next assessed the effect of the rs4844609^A allele on clinically diagnosed AD. In our ROS and MAP
discovery cohorts, we observed a two-fold increase in the odds of clinical AD associated with rs4844609A (Odds Ratio (OR) = 2.2, p=0.007; 95% Confidence Interval (CI): 1.25-3.87) (Table 3). This association remained significant (OR = 2.0, p=0.028; 95% CI: 1.08-3.81) after adjusting for rs6656401. Interestingly, no significant associations were observed between clinically diagnosed AD and rs6656401 (p=0.104 in independent model; p=0.582 after adjusting for rs4844609) in our discovery sample, which is of modest size and is underpowered to replicate the effect of the index SNP on AD susceptibility. The results of our replication analyses for AD susceptibility are presented in SI Table S5. Replication efforts are limited by sample size, study design, and quality of the imputation of rs4844609 allele dosage in each stratum of the meta-analysis. While two of the three strata in the replication meta-analysis demonstrate a direction of effect consistent with the discovery cohorts, the forest plot illustrates the non-significant trend towards replication (SI Figure S3). A joint analysis of all data remains significant and suggests a strong effect on AD susceptibility (OR 1.40, 95% CI: 1.02-1.94) by rs4844609A; however, more subjects and direct genotyping of the variant will be needed to provide a definitive assessment of the role of rs4844609 in AD.

**Haplotype Analysis.** Another approach to assessing the relative effects of the rs4844609 and rs6656401 variants involves implementing a haplotype-based analysis (Figure 5). Among our ROS and MAP subjects, we observed 12 haplotypes of >1% frequency in this region. The risk allele of the index SNP (rs6656401A) is found in three of these haplotypes, one of which is the only haplotype containing rs4844609A (frequency=0.02). The haplotype containing both risk alleles is the only one significantly associated with clinical AD in our data (p=0.001 in univariate analyses, as covariates cannot be introduced in Haplovview). In order to conduct a conditional
haplotype analysis, we note that our coding variant rs4844609 distinguishes this risk haplotype from the other 11 haplotypes in our sample. Therefore, analyzing rs4844609 as a binary variable (any or none) is equivalent to assessing this risk haplotype against all others. When we perform this analysis, we observe significant associations with clinically diagnosed AD (OR=2.09, p=0.006; 95% CI: 1.24-3.53), episodic memory decline ($\beta$(SE)=-0.034(0.012), p=0.005), and neuritic plaque burden ($\beta$(SE)=0.244(0.106), p=0.021), adjusted for age, sex, education, cohort, and the first 3 principal components generated from genome-wide genotype data that capture the effect of sub-structure in European ancestry amongst our subjects. This analysis suggests that, of the considered polymorphisms, rs4844609 best captures the effect of the CR1 locus on AD susceptibility in our data, and that rs6656401 may be a surrogate marker harbored on the same haplotype ($D'=1$ and $r^2=0.084$ between rs4844609 and rs6656401). The more frequent (15%) haplotype containing only the risk allele of the index SNP demonstrates no association with AD susceptibility in our data (p=0.829, Figure 5).

Many polymorphisms relating to the function of CR1 have been described, in particular the CR1-S and CR1-F alleles that are defined on the basis of the mobility of CR1 in gel electrophoresis experiments (S=slow; F=fast). This test captures the fact that the CR1-S polymorphism has an extra copy of the long low-copy repeat 1 (LCR1), which results in an additional LHR domain (LHR-S) and, consequently, additional C3b/C4b binding sites (Figure 1). Brouwers and colleagues (17) recently reported that the CR1-S allele, with a MAF of 0.15 in their study, is in strong LD with the index CR1 SNP associated with susceptibility to AD and is therefore found on the haplotype that has been associated with AD susceptibility. Brouwers and colleagues focus more closely on two variants that are associated with AD risk, rs4844610 and rs1408077, which are in strong LD with both the copy number variant and rs6656401. To
provide a more directed assessment of the relative roles of the CR1-S allele and the rs4844609 coding variant, we also performed a haplotype analysis using only these 4 SNPs (rs6656401, rs4844609, rs4844610, and rs1408077). This analysis returns evidence for 6 distinct haplotypes in our population (SI Table S6), 3 of which have a frequency of greater than 1%. Of these haplotypes, only the haplotype containing the risk allele of our coding variant (rs6656401^A-rs4844609^A-rs4844610^A-rs1408077^A, frequency=0.02) is significantly associated with AD susceptibility (p=0.001). This haplotype contains the CR1-S allele and the risk allele of our coding variant, as well as the risk allele of the index SNP, rs6656401. We do not observe a significant association between the most frequent (15%) haplotype which contains the index risk allele, the CR1-S allele but not the coding variant (rs6656401^A-rs4844609^T-rs4844610^A-rs1408077^A, p=0.504). We therefore refine the earlier report on the possible role of CR1-S (17) and suggest that the coding variant rs4844609^A may be found within the context of the larger CR1 molecule which contains the additional LHR domain that is encoded in the CR1-S polymorphism.

**Exploring the functional consequences of rs4844609^A.** Given our prior work suggesting that CR1 was associated with amyloid plaque burden in a subset of 553 of the currently available sample of 651 autopsies (12), we next examined the association of rs4844609^A with neuritic amyloid plaques and found that rs4844609^A was also associated with increased neuritic pathology (p=0.011, Table 3). The effect size we observe for our coding variant is larger than the one previously observed for the index SNP (12). Joint models including both SNPs once again suggest that rs4844609^A best captures the association of this trait with the CR1 locus (Table 3). Further, we also see evidence for an increased burden of neurofibrillar tangles in subjects
bearing the rs4844609A allele (p=0.021), and the association remains significant after adjusting for rs6656401 (Table 3). As previously reported, we do not observe evidence of association between rs6656401 and neurofibrillary tangle burden in our subjects (12).

We next evaluated whether these associations with the burden of neuritic plaques and neurofibrillary tangles mediate the effect of our CRI coding variant on cognitive decline. Adjusting for the burden of neuritic plaques reduces the observed effect of rs4844609A on episodic memory by 35% [from \( \beta=-0.039 \) (p=0.003) to \( \beta=-0.026 \) (p=0.273)], and adjusting for neurofibrillary tangles reduces the effect by 41% [from \( \beta=-0.039 \) (p=0.003) to \( \beta=-0.023 \) (p=0.309)]. Thus, we extend our earlier observation that the effect of the CRI locus on episodic memory decline is mediated, in part, by an effect on amyloid pathology: we now also have evidence implicating Tau pathology in the effect of the CRI locus. Since both pathologies are closely correlated, additional studies using molecularly specific markers will be needed to distinguish their respective effects. Interestingly, both pathologies are also influenced by APOE-\( \varepsilon4 \) and mediate its effect on cognitive decline (18-20), suggesting a potential molecular correlate to the evidence of statistical interaction that we report between the CRI and APOE susceptibility alleles.

**Discussion**

Recent studies have found strong evidence for associations between a CRI SNP, rs6656401, and both clinically diagnosed AD and aging-related cognitive decline (2, 4, 12). In this manuscript, we sought to identify variants within the CRI locus that have functional consequences on traits related to AD. We report evidence supporting the role of a non-synonymous coding variant in the LHR-D region of the CRI gene (rs4844609) that is strongly associated with episodic memory
decline; further, rs4844609 accounts for the previously observed effect of rs6656401 on this intermediate phenotype. In secondary analyses (SI Table S7), we see the same relationships with another intermediate phenotype, global cognitive decline, which incorporates our measures of episodic memory along with assessments of four other cognitive domains. Looking at these other domains separately, we also observe associations with perceptual speed (p=0.025) and semantic memory (p=0.015); thus the effect of rs4844609 does not appear to be limited to decline in episodic memory. The rs4844609 variant is a non-synonymous coding SNP found in exon 37 of CR1, which encodes a portion of its LHR-D domain. Taken together, our results suggest that we have found a functional variant that accounts for the effect of the CR1 locus on decline in episodic memory. Our analyses also suggest that this variant may influence AD susceptibility, but this association remains to be further delineated in larger, well-powered studies.

CR1 binds all of the nascent complement opsonins, i.e. C4b, C3b, mannan-binding lectin (MBL), and C1q. C4b, C3b, and MBL have roles in the clearance of microbes, but C1q has a unique role in the removal of apoptotic/necrotic debris: greater than 90% of subjects with C1q deficiency suffer a severe lupus-like illness with onset in childhood and a high morbidity and mortality (21). CR1 is a critical component of the complement receptor pathway that serves as a scaffold for different complement molecules in its LHR-A, -S, -B, and -C domains, but also mediates the removal of opsonized material or cells by binding the opsonins C1q and MBL in its LHR-D domain, which contains the rs4844609A allele associated with cognitive decline (11, 22, 23). The exact effect of this allele, which results in a Threonine for Serine substitution at position 1610, on the function of the LHR-D domain in CR1 remains unclear. Nonetheless, it alters the structure of CR1 sufficiently to cause an immune response and anti-CR1 antibody formation.
following blood transfusion in humans (24), suggesting that the conformational change is substantial.

The identification of the functional unit of CR1 that is involved in decline in episodic memory is an important step forward, implicating several other molecules in the effect of rs4844609A. We speculate that this variant may reduce the binding affinity of the cognate LHR-D ligands, C1q and MBL, reducing the efficacy of the clearance of opsonized material such as amyloid oligomers or amyloid plaques. While little is known about MBL in AD aside from its decreased level in cerebrospinal fluid of AD subjects (25), C1q is a complement molecule that has been studied extensively in the context of AD. Specifically, in humans, C1q has been reported to be 1) more abundant in AD brains, particularly in brain regions with dense accumulations of amyloid plaques and neurofibrillar tangles (26), 2) associated with fibrillar Aβ and amyloid plaques (27-30), and 3) found on certain neurons in AD brains that are associated with microglia (31). Further, in different mouse models of AD, C1q is transcriptionally upregulated during aging and in the context of amyloid deposition (32, 33), and it is found to be associated with fibrillar Aβ (34). In addition, C1q null animals, in some transgenic models, have less accumulation of amyloid plaques (28, 35). Thus, while the exact role of C1q in the context of AD needs to be further delineated, human tissue and rodent model systems implicate it in the pathophysiology of AD and provide a rich framework from which to design studies to investigate the role of the interaction of CR1’s LHR-D domain with C1q in the clearance of amyloid plaques. While we speculate that clearance of opsonized plaque material may be involved, we note that CR1 ligation also causes, in nucleated cells, phosphorylation of CR1, Ca²⁺ influx and phospholipase D activation, which may trigger immunomodulatory responses by B cells and possibly other immune cells (10, 36, 37). Consequently, our coding variant directly connects a
gene influencing decline in episodic memory to a rich literature that implicates opsonization in the biology of amyloid plaques in the aging brain.

Extending the prior observation of interaction between these two loci in terms of AD susceptibility (4), our data suggest that the effect of our coding CR1 variant on cognitive decline may be dependent on the presence of the APOE-ε4 allele. The confluence of these results with the observation that APOE-ε4 also affects both amyloid and tau pathologies in our cohorts (18-20) suggests that these two molecules may be affecting a single molecular process that has a large effect on susceptibility to AD.

One important limitation for our current analyses is that data for the coding variant is imputed: we use allele dosage files in our analyses in order to mitigate the uncertainty that is inherent to imputation techniques. While the quality of this imputation varies among our discovery and replication datasets, we note that poor imputation quality does not overestimate our associations but actually decreases our power and biases our results towards the null. Moreover, the authors recognize that it is important to interpret associations with an imputed genotype of low frequency cautiously. To address this issue, we performed direct genotyping of the rs4844609 and rs6656401 variants in a subset of subjects with available DNA, obtaining genotyping information on 1594 (93% of the sample with genome wide imputed data) after quality control (QC). Within this sample, we found a strong correlation between the imputed dosages and actual genotypes ($r^2=0.814$, $p<0.0001$). Of these subjects, information on episodic memory was available on 1485 (93%) and 318 (20%) were clinically diagnosed with AD. The mean (SD) age at baseline in this sample was 78.5 (7.5) and 495 (31%) were male. We observed associations between these primary phenotypes and the genotyped SNPs that are consistent with our results based on the imputed dosages (SI Table S8). Finally, since the original submission of
the manuscript, imputed dosage data using a different reference map, the 1000 Genomes reference panel (dated November 23, 2010 and downloaded in July 2011) (38), became available in our subjects, and 131 additional SNPs can now be interrogated within the CR1 LD block. However, none of the newly imputed SNPs provide more extreme evidence of association to our outcome measures than we have observed with rs4844609 (SI Figure S4). Ultimately, resequencing of the CR1 locus in a large number of subjects will be required to directly catalog all genetic variation in this genomic segment and to definitively identify the causal variant(s) affecting CR1 function.

Another limitation of these analyses is that our population consists of highly educated individuals of European ancestry, and thus we should be careful about generalizing our results to other populations. Nonetheless, many aspects of our study design can be viewed as strengths. Subjects in both ROS and MAP have high rates of both follow-up and autopsy, limiting the opportunity for bias. Our analyses relied on structured clinical and pathologic procedures and quantitative measures of AD pathology, which have excellent metric properties and increase both our ability and power to discover functional pathways associated with AD.

Overall, our results suggest that we have found a coding variant in the CR1 locus that influences decline in episodic memory, and that this allele’s effect on decline in episodic memory may be mediated in large part by enhancing the accumulation of amyloid and tau pathologies. The path forward is clear: future studies are needed to explore the role of this variant in AD susceptibility and to examine the functional consequences of rs4844609A in order to explore the hypothesis that this allele, in the context of APOE-ε4, alters responses to AD pathology. With the identification of the rs4844609A allele, manipulation of exon 37 in CR1 or, more broadly, of the function of the LHR-D domain is clearly suggested as an approach to
understand the role of $CR1$ in amyloid and tau-mediated pathologies that are found in AD.

**Materials and Methods**

**Discovery Sample.** The primary analyses used data from participants from two cohort studies, the *Religious Orders Study* (ROS) and *Rush Memory and Aging Project* (MAP), described in detail in the SI Methods and published literature (39, 40). Briefly, ROS was established in 1994 and consists of Catholic clergy from the United States aged 55 or older and free of known dementia at enrollment. Since January 1994, 1100 participants (1001 self-reported non-Hispanic white) completed their baseline evaluation, and follow-up and autopsy rate exceed 90%. MAP was established in 1997 and consists of older men and women from continuous care retirement centers in the Chicagoland area, aged >55 and free of known dementia. Since October 1997, 1400 participants (1118 self-reported non-Hispanic white) completed their baseline questionnaire; follow-up rate exceeds 90% and autopsy rate exceeds 80%. As in prior studies (12, 13, 39), we analyzed ROS and MAP subjects jointly, adjusting for possible confounders, and since 1) they were designed to be combined, 2) the outcome measures and data collection procedures required for these analyses are identical, and 3) they are managed by a single investigative team. Both studies were approved by the Institutional Review Board of Rush University Medical Center.

**Replication Samples.** Our Replication samples consisted of two prospectively followed samples, the *Alzheimer’s Disease Neuroimaging Initiative* (ADNI) and the *Chicago Health and Aging Project* (CHAP), and one large case-control sample from the *Translational Genetics*
Research Institute (TGEN). Detailed descriptions of these cohorts are presented in the SI Methods.

**Phenotypic data.** Our primary measure of cognition was episodic memory, a composite score of 7 instruments in ROS and MAP (word list memory, recognition and recall from CERAD, immediate and delayed recall of Story A from the logical memory subset of the Wechsler Memory Scale-Revised, and immediate and delayed recall of the East Boston Story), 8 instruments in ADNI (word list memory, recognition and recall from CERAD, immediate and delayed recall of Story A from the logical memory subset of the Wechsler Memory Scale-Revised, and the Rey Auditory Verbal Learning Trials 1-6 totals, 30 minute delay total, and recognition score), and 2 instruments in CHAP (immediate and delayed recall of the East Boston Story). For each cohort, we used previously described methods (12, 40-42) to create a normalized summary measure relative to the baseline population. General linear mixed models were used to model episodic memory decline over time within each cohort, adjusted for age at enrollment, years of education, and sex. From these models, we obtained estimated residual slopes for each individual with at least 2 recorded measures of episodic memory.

Clinical diagnosis of probable AD was based on the NINCDS-ADRDA criteria (43) in ROS, MAP, ADNI, and CHAP, or clinically characterized and neuropathologically defined in TGEN. We created standardized summary measures of neurofibrillary tangles and neuritic plaques in ROS and MAP by dividing each person’s count by the standard deviation for that particular count and taking the square root of this value to normalize the outcome, as previously reported (13, 20). More phenotype information is presented in the SI methods and previously published literature (12, 39, 40, 44-46).
Genotypic data. Genetic information included genotyped and imputed SNPs. Detailed genotyping and quality control methods for the discovery and replication cohorts are described in detail in the SI Methods and elsewhere (2, 12, 47-49). Imputation in all cohorts was performed using MACH software (version 1.0.16a) and HapMap release 22 CEU (build 36) or the 1000 Genomes Project as a reference (38). The $R^2$ quality measure for the index $CR1$ SNP rs6656401 was 0.9997 in ROS and MAP, 0.9270 in ADNI, 0.9091 in CHAP, and 0.9890 in TGEN. The $R^2$ for the non-synonymous coding SNP rs4844609 was 0.7212 in ROS and MAP, 0.1426 in ADNI, 0.6868 in CHAP, and 0.5570 in TGEN.

Statistical methods. Detailed statistical methods are presented in the SI Methods. Briefly, within each cohort, general linear mixed models were used to obtain residual slope estimates of episodic memory decline for each individual with at least 2 recorded measures. Using these estimates as outcomes, we fit linear regression models within our discovery cohort, examining associations with the SNPs in the $CR1$ LD block. Candidate functional variants were those SNPs that were highly associated with episodic memory and had evidence of functional significance. For secondary AD outcomes, we examined the association with candidate functional variants in each sample using linear regression for continuous AD pathology measures and logistic regression for AD diagnoses. Where appropriate, we conducted a fixed effects meta-analysis across the replication datasets and a joint meta-analysis incorporating ROS and MAP. Models in ROS and MAP, ADNI, and CHAP were adjusted for the first three principal components, age at enrollment, years of education, and sex. Models in ROS and MAP were additionally adjusted for cohort. Models in TGEN were adjusted for sex and age at death. To determine whether the
candidate causal variant accounted for the observed effect of the index *CR1* SNP, we additionally included both SNPs as covariates, and examined the changes in effect sizes compared to the single SNP models. To assess our hypothesized causal chains of intermediate phenotypes, we further adjusted for AD pathology outcomes associated with our functional variants of interest. We assessed whether there was evidence of interaction between *APOE*-ε4 and identified functional variants by including both the main effects for each variant and a product term (*APOE*-ε4 x SNP) in the regression models. If we observed a p<0.05 for interaction, we further examined the effect of our candidate functional variant in stratified subsets. Statistical analyses were performed using SAS software, version 9.1 (SAS Institute, Cary, NC), R, version 2.13 (www.r-project.org) and the PLINK toolkit (http://pngu.mgh.harvard.edu/~purcell/plink/). The meta-analysis was conducted using R, version 2.13 (www.r-project.org) and METAL (50).

Acknowledgements

The authors are grateful to the participants in the Religious Orders Study, the Rush Memory and Aging Project, the Alzheimer's Disease Neuroimaging Initiative, the Chicago Health and Aging Project, and the Translational Genetics Research Institute, and thank Alessandro Biffi, Christopher Anderson and Jonathan Rosand for sharing the quality-controlled genotyping dataset from ADNI.

This work is supported by the National Institutes of Health (NIH) [R01 AG30146, R01 AG179917, R01 AG15819, K08 AG034290, P30 AG10161 and R01 AG11101] and the Illinois Department of Public Health. MJH, JC and EMR are supported by grants from the NIH-NINDS [R01NS059873], Kronos Life Sciences Laboratories, the National Institute on Aging (NIA)
(Arizona Alzheimer’s Disease Center P30 AG19610, R01 AG023193, Mayo Clinic Alzheimer’s Disease Center P50 AG16574 and Intramural Research Program), the National Alzheimer’s Coordinating Center (U01 AG016976), the NIH Neuroscience Blueprint (U24NS051872), the ENDGAME Consortium (UO1HL084744), NIA grants to Carl Cotman (University of California, Irvine, P50 AG23173), and K01AG024079, and the state of Arizona. AJM is supported by NIA grant R01 AG034504 and the Johnnie B. Byrd Institute. JAH would like to thank the Verum foundation. Data collection and sharing for ADNI was funded by the NIH Grant U01 AG024904. ADNI is funded by the NIA, the National Institute of Biomedical Imaging and Bioengineering, through generous contributions 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 Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., and non-profit partners the Alzheimer’s Association and Alzheimer’s Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions are facilitated by the Foundation for the NIH (www.fnih.org). The grantee organization is the Northern 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 Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest Statement

The authors have no conflicts of interest in regard to this manuscript.
References


Figure Legends

Figure 1. Schematic representation of CCPs, LHRs, and binding sites of the most common CR1 molecule. The most common form of CR1, containing 30 complement control repeats (CCPs), is depicted. The first 28 CCPs form 4 long homologous repeat domains (LHR A-D), shown above the CR1 molecule. The areas of each LHR containing likely C4b, C3b, MBL and/or C1q binding sites are illustrated below the molecule. Also shown above the molecule is the location and structure of the LHR-S polymorphism described by Brouwers et al (17). NH2 represents the amino terminus, TM the transmembrane domain, and CYT the cytoplasmic tail. Finally, the red star denotes the location of the functional variant rs4844609. Schematic derived from similar figures in Krych-Goldberg et al (8), Liu et al (9) and Brouwers et al (17).

Figure 2. Association plot for decline in episodic memory in the CR1 region. Each square represents the evidence of association for a particular SNP, the physical position of which is presented on the x-axis. The intensity of red color, presented graphically in the top right corner, represents the LD between each SNP and rs6656401. The blue line reports the recombination rate. The plot highlighting rs4844609 is presented in SI Figure S1.

Figure 3. Episodic memory decline in ROS and MAP subjects, stratified by CR1 genotypes. The effect of the rs6656401A and rs4844609A risk alleles on episodic memory decline is shown by computing the mean trajectory of decline for the different genotype categories. The graph presents the extent of change in episodic memory from baseline (at enrollment). The 2x2 table depicted shows both the sample size and the color of the line representing the slope of episodic memory decline for that genotype group. There are no rs4844609AA homozygotes in our dataset. The groups are stratified by presence of both rs6656401 and rs4844609 [p=0.013 comparing rs6656401GG + rs4844609TT to rs6656401AT/AA + rs4844609AT; p=0.052 comparing rs6656401GG + rs4844609TT to rs6656401AT/AA + rs4844609TT; p=0.1253 comparing rs6656401AG/AA + rs4844609TT to rs6656401AT/AA + rs4844609AT].

Figure 4. Episodic memory decline in ROS and MAP subjects, stratified by APOE-ε4 and rs4844609. The slope of episodic memory decline for the four distinct subsets defined by the presence or absence of APOE-ε4 and rs4844609A risk alleles is illustrated. The 2x2 table depicted shows both the sample size and the color of the line representing the slope of episodic memory decline for each of the 4 strata. Subjects with both rs4844609AT and APOE-ε4 decline significantly faster than all other groups [p<0.0001 compared to rs4844609TT + no APOE-ε4; p<0.0001 compared to rs4844609AT + no APOE-ε4; p=0.005 compared to rs4844609TT + APOE-ε4].

Figure 5. Haplotypes and associations with clinical AD in the CR1 gene. Using the “solid spine of LD” method of defining LD blocks in the CR1 region, we identified 12 haplotypes of greater than 1% frequency among ROS and MAP subjects, spanning most of the CR1 gene (shown in purple). The ancestral haplotype (derived from the chimp) is shown at the top in black. The 12 haplotypes and frequencies in our sample are shown below the ancestral haplotype, with the derived alleles highlighted in purple and marked with black dots. The presence of the rs6656401A and rs4844609A risk alleles is marked with red dots. The index CR1 risk allele appears on 3 haplotypes, only one of which contains our candidate causal variant. The putative phylogenetic tree is rooted in the ancestral haplotype and illustrates a possible evolutionary relationship of the different haplotypes. P-values for crude associations with clinically diagnosed AD‡ are shown on the right.
Table 1. Associations with episodic memory decline for CR1 SNPs in ROS and MAP* 

<table>
<thead>
<tr>
<th></th>
<th>Independent Models</th>
<th>Joint Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β (SE) p</td>
<td>β (SE) p</td>
</tr>
<tr>
<td>rs6656401^A</td>
<td>-0.009 (0.005) 0.048</td>
<td>---</td>
</tr>
<tr>
<td>rs4844609^A</td>
<td>---</td>
<td>-0.039 (0.013) 0.003</td>
</tr>
</tbody>
</table>

Combined ROS and MAP cohort; n=1593 subjects with non-missing values; SE=standard error

Table 2. Replication analyses with episodic memory decline for CR1 SNPs 

<table>
<thead>
<tr>
<th></th>
<th>ADNI (N=716)</th>
<th>CHAP (N=569)</th>
<th>Meta-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β (SE) p</td>
<td>β (SE) p</td>
<td>p</td>
</tr>
<tr>
<td>rs6656401^A</td>
<td>-0.001 (0.001) 0.222</td>
<td>-0.002 (0.003) 0.618</td>
<td>0.162</td>
</tr>
<tr>
<td>rs4844609^A</td>
<td>-0.007 (0.003) 0.014</td>
<td>-0.016 (0.010) 0.100</td>
<td>0.004</td>
</tr>
</tbody>
</table>

p-value from fixed effects meta-analysis adjusted for sample size and direction of effect; SE= standard error

Table 3. Associations with secondary phenotypes for CR1 SNPs in ROS and MAP Cohort 

<table>
<thead>
<tr>
<th></th>
<th>Independent Models</th>
<th>Joint Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clinically Defined AD*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR (95% CI) p</td>
<td>OR (95% CI) p</td>
</tr>
<tr>
<td>rs6656401^A</td>
<td>1.20 (0.96,1.50) 0.104</td>
<td>---</td>
</tr>
<tr>
<td>rs4844609^A</td>
<td>---</td>
<td>2.19 (1.25,3.87) 0.007</td>
</tr>
</tbody>
</table>

Neuritic Plaques†

<table>
<thead>
<tr>
<th></th>
<th>β (SE) p</th>
<th>β (SE) p</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6656401^A</td>
<td>0.084 (0.040) 0.035</td>
<td>---</td>
<td>0.052 (0.043) 0.231</td>
</tr>
<tr>
<td>rs4844609^A</td>
<td>---</td>
<td>0.286 (0.113) 0.011</td>
<td>0.227 (0.123) 0.065</td>
</tr>
</tbody>
</table>

Neurofibrillary Tangles‡

<table>
<thead>
<tr>
<th></th>
<th>β (SE) p</th>
<th>β (SE) p</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6656401^A</td>
<td>0.027 (0.029) 0.365</td>
<td>---</td>
<td>-0.0002 (0.032) 0.994</td>
</tr>
<tr>
<td>rs4844609^A</td>
<td>---</td>
<td>0.192 (0.083) 0.021</td>
<td>0.192 (0.091) 0.035</td>
</tr>
</tbody>
</table>

340 Cases and 1368 Controls, effect presented as OR (95% CI); †651 deceased subjects with non-missing measures, effect presented as β (standard error); SE = standard error; OR = Odds Ratio
Abbreviations

AD = Alzheimer’s Disease
GWAS = Genome wide association study
APOE = Apolipoprotein E
CR1 = Complement Receptor 1
CCP(s) = complement control repeat(s)
LHR(s) = long homologous repeat(s)
SNP(s) = Single Nucleotide Polymorphism(s)
ROS = Religious Orders Study
MAP = Rush Memory and Aging Project
SI = Supporting Information
ADNI = Alzheimer’s Disease Neuroimaging Initiative
CHAP = Chicago Health and Aging Project
SD = Standard Deviation
TGEN = Translational Genetics Research Institute
LD = Linkage Disequilibrium
MAF = Minor Allele Frequency
nscSNP = non-synonymous coding single nucleotide polymorphism
OR = Odds Ratio
CI = Confidence Interval
LCR1 = low-copy repeat 1
MBL = mannan-binding lectin
QC = quality control
NINCDS-ADRDA = National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association
SE = standard error
Fig. 2
Fig. 3
Fig. 4