Replication of *CLU, CR1* and *PICALM* associations with Alzheimer's disease

Minerva M. Carrasquillo[1], Olivia Belbin[1], Talisha A. Hunter[1], Li Ma[1], Gina D. Bisceglio[1], Fanggeng Zou[1], Julia E. Crook[2], V. Shane Pankratz[3], Dennis W. Dickson[1], Neill R. Graff-Radford[1,4], Ronald C. Petersen[4,5], Kevin Morgan[6] and Steven G. Younkin[1]

[1]Department of Neuroscience, and [2]Biostatistics Unit, Mayo Clinic College of Medicine, Jacksonville, FL 32224, USA, [3]Division of Biomedical Statistics and Informatics, [4]Department of Neurology, and [5]Mayo Alzheimer Disease Research Center, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA, [6]School of Molecular Medical Sciences, Institute of Genetics, Queens's Medical Centre, University of Nottingham, UK

Correspondence should be addressed to: younkin.steven@mayo.edu, Mayo Clinic, 4500 San Pablo Rd, Jacskonville, FL 32224, telephone: (904) 953-7356, fax: (904) 953-73570

Revised: 01-13-10

Word Count (for text, inclusive of abstract, references, tables, and figure legends): 2,856

Abstract

Background: Harold et al. and Lambert et al. recently published two large genome-wide association studies of late onset Alzheimer's disease (LOAD) in which *CLU*, *CR1*, and *PICALM* were identified as novel LOAD genes.

Objective: To test for replication of the association between variants in the *CLU*, *CR1* and *PICALM* genes with Alzheimer's disease.

Design: Case-control association study

Setting: Community-based ascertainment of patients seen at the Mayo Clinic Jacksonville, FL and Rochester, MN, and autopsy-confirmed cases and controls whose pathology was evaluated at the Mayo Clinic Jacksonville. Additional samples were obtained from the National Cell Repository for Alzheimer's Disease (NCRAD).

Participants: LOAD case-control series of European descent consisting of 1,829 LOAD cases and 2,576 controls

Main Outcome Measure: Clinical or pathology-confirmed diagnosis of LOAD

Results: In our follow-up study of 1,829 LOAD cases and 2,576 controls, the most significant SNPs in *CLU* (rs11136000), *CR1* (rs3818361), and *PICALM* (rs3851179)

were tested for allelic association and gave ORs of 0.82, 1.15, and 0.80 respectively that were comparable in direction and magnitude to those originally reported with p values of $8.6x10^{-5}$, 0.014, and $1.3x10^{-5}$ that were significant even after Bonferroni correction for 3 SNPs tested.

Conclusion: These results showing near perfect replication provide the first additional evidence that *CLU*, *CR1*, and *PICALM* are LOAD genes.

Introduction

Late onset Alzheimer's disease, a neurodegenerative condition characterized by large numbers of senile plaques and neurofibrillary tangles in the brain, is the most common cause of dementia in the elderly. Multiple rare mutations in the *APP*, *PSEN1* and *PSEN2* genes cause an early onset familial form of the disease (for review see ref. [1]), and twin studies indicate that susceptibility alleles may contribute as much as 80% to the risk of late-onset AD (LOAD).[2] Until recently, however, *APOE ε4* was the only allele reliably associated with increased susceptibility to LOAD.[3-5] A robust technology has emerged that permits genome-wide association studies (GWAS) of large numbers of subjects. This technology has enabled the identification of relatively weak associations that would otherwise go undetected.

Recently Harold *et al.*[6] and Lambert *et al.*[7] published the two largest LOAD GWAS conducted to date and reported genome-wide significant association with three novel LOAD genes. The first study, by Harold *et al.*, reported association of single nucleotide polymorphisms (SNPs) in *CLU* and *PICALM.* The second study, by Lambert *et al.*, also reported association of *CLU* with LOAD and additionally reported novel association with *CR1*. Here we report our effort to replicate these findings in an independent LOAD case-control series of European descent consisting of 1,829 LOAD cases and 2,576 controls. Our results show near perfect replication and provide the first additional evidence for association of these three genes with LOAD.

Methods

Case-control subjects

Samples used in this study do not overlap with the samples included in the Harold *et al.*[6] or Lambert *et al.*[7] publications. The USA case-control series consisted of Caucasian subjects from the United States ascertained at the Mayo Clinic or through the National Cell Repository for Alzheimer's Disease (NCRAD). All subjects ascertained at the Mayo Clinic in Jacksonville, Florida (JS) and at the Mayo Clinic in Rochester, Minnesota, (RS) were diagnosed by a Mayo Clinic neurologist. The neurologist confirmed a Clinical Dementia Rating score of 0 for all JS and RS subjects enrolled as controls; cases had diagnoses of possible or probable AD made according to NINCDS-ADRDA criteria.[8] In the autopsy-confirmed series (AUT), all brains were evaluated by Dr. Dennis Dickson and came from the brain bank maintained at the Mayo Clinic in Jacksonville, FL. In the AUT series the diagnosis of definite AD was also made according to NINCDS-ADRDA criteria. All AD brains analyzed in the study had a Braak score of 4.0 or greater. Brains employed as controls had a Braak score of 2.5 or lower but often had brain pathology unrelated to AD and pathological diagnoses that included vascular dementia, frontotemporal dementia, dementia with Lewy bodies, multi-system atrophy, amyotrophic lateral sclerosis, and progressive supranuclear palsy. One AD case from each of the 702 late-onset NCRAD families was analyzed. NCRAD AD cases were selected based on strength of diagnosis (autopsy-confirmed: 32% > probable: 45% > possible: 8% > family report: 15%); the case with the earliest age at diagnosis was taken when several cases had equally strong diagnoses. The 209 NCRAD controls that we

employed were unrelated Caucasian subjects from the United States with a Clinical Dementia Rating of 0, specifically collected for inclusion in case-control series. Written informed consent was obtained for all individuals that participated in this study.

DNA isolation

For the JS and RS samples, DNA was isolated from whole blood using an AutoGen instrument (AutoGen, Inc, Holliston, MA). The DNA from AUT samples was extracted from cerebellum using Wizard® Genomic DNA Purification Kits (Promega Corp., Madison, WI). DNA from the RS and AUT series was scarce, so samples from these two series were subjected to whole genome amplification using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences Corp., Piscataway, NJ).

Genotyping

All genotyping was performed at the Mayo Clinic in Jacksonville using TaqMan® SNP Genotyping Assays in an ABI PRISM® 7900HT Sequence Detection System with 384-Well Block Module from Applied Biosystems, California, USA. The genotype data was analyzed using the SDS software version 2.2.3 (Applied Biosystems, California, USA).

Statistical Analyses

For the analysis of the Mayo series, *a*n allelic dosage model using logistic regression adjusted by sex and *APOE ε4* status was used to generate p values and odds ratios for the association of LOAD with the minor allele. The results reported in the two original GWAS papers by Harold *et al*.[6] and Lambert *et al*.[7] were also generated using

logistic regression analyses assuming an additive model to test for association with the minor allele. However, the covariates used differed between the three studies. Harold *et al*. did not adjust for age or gender, only for geographical region and genotyping array used; while Lambert *et al*. adjusted for age, gender and center. The results from the Mayo series stay essentially the same when no covariates are used (i.e. when no adjustment is made for sex and *APOE ε4* status. Since genotype counts were not available for the Harold *et al*. follow-up series, or for either the GWAS or follow-up series from the Lambert *et al*. study, we employed a Fisher's combined test to combine p values across series.

Results

We genotyped the most significant SNP in each of the three genes reported by Harold *et al*.[6], and Lambert *et al*.[7], to determine if the associations could be detected in our follow-up series. Remarkably, the direction and magnitude of each association replicated well in our series and addition of our follow-up data to the results previously reported increased the strength of evidence (Fisher's combined summary statistic) for each of the associations (Table 1).

The Harold *et al*. GWAS showed significant association of the *CLU* SNP, rs11136000, with reduced risk of LOAD (OR = 0.86, 95% CI 0.82-0.90, $p = 8.5 \times 10^{-10}$) in their combined series (GWAS + "extension series", 5,964 LOAD cases and 10,188 controls). This association was replicated in the Lambert *et al*. study (OR = 0.86, 95% CI 0.81-0.90, $p = 7.5 \times 10^{-9}$) in their combined series (5,791 LOAD cases and 8,420 controls).

In our LOAD case-control follow-up series of 1,819 LOAD cases and 2,565 controls, we observed an odds ratio of the same magnitude and direction (OR = 0.82, 95% CI 0.75-0.91, $p = 8.6x10^{-5}$).

In their GWAS, Lambert *et al*. reported a significant association of the *CR1* SNP, rs3818361, with increased risk of LOAD (OR = 1.19, 95% CI 1.11-1.26, $p = 8.9x10^{-8}$), an effect that was replicated by Harold *et al*. in their GWAS series of 3,939 LOAD cases and 7,848 controls (OR = 1.17, 95% CI 1.09-1.25, $p = 9.2x10^{-6}$). Again, we observed an odds ratio of the same magnitude and direction (OR = 1.15, 95% CI 1.03-1.29, $p = 1.4x10^{-2}$).

Finally, Harold *et al*., reported significant association of the *PICALM* SNP, rs3851179, with reduced risk of LOAD (OR = 0.86, 95% CI 0.82-0.90, $p = 1.3x10^{-9}$). This association was not reported by Lambert *et al*., but we observed the same association as Harold *et al*., in our follow-up study (OR = 0.80, 95% CI 0.73-0.89, $p = 1.3x10^{-5}$).

Comment

For over 15 years, *APOE* alleles were the only genetic variants that showed replicable association with altered susceptibility to LOAD. Over 500 additional candidate genes were investigated in more than 1,200 studies,[9] but little progress was made until the last several years when large LOAD case-control series and meta-analyses were employed to gain the power necessary to detect associations much weaker than those of the *APOE* alleles. Except for the three genes pursued in this manuscript, the most

significant evidence for a novel LOAD gene comes from the two-stage GWAS of Carrasquillo et al.[10], who found impressive association of *PCDH11X* (rs2573905) with LOAD ($p = 5.4x10^{-13}$) in their two-stage GWAS of seven case-control series with a combined total of 5,010 subjects. The AlzGene website maintained by Bertram and Tanzi, which summarizes genetic association studies of LOAD, currently lists 35 loci (as of November 2009) with one or several variants that are nominally significant when tested for allelic association in random-effects meta-analyses.[11] Some of these were identified using a candidate gene approach, others in the eleven LOAD GWAS (e.g. *CLU*, *CR1*, *PICALM*, and *GAB2*) that have been performed to date (for review see Bertram and Tanzi.[9]) *SORL1*, which directs Aß away from the Aß-generating pathway and was discovered using a candidate gene approach,[12] shows particularly impressive association with several variants that associate with LOAD in multiple, large case-control series. The association for many of the other candidate genes is based on relatively few subjects, shows substantial heterogeneity from series, or is weak. Thus many of the nominally significant associations for candidate genes are tenuous and require additional replication.

Genome-wide association studies often give inflated ORs that are substantially reduced in follow-up series. In this first, independent follow-up analysis of *CLU*, *CR1*, and *PICALM*, this was not the case, as we obtained ORs that were essentially identical to those observed in the initial studies. Thus our findings provide strong additional evidence that all three genes are novel LOAD genes. *CLU* (aka *APOJ*) encodes clusterin which interacts with Aß[13-16] and appears to influence the aggregation and toxicity of this important AD-related peptide.[17-20] *CR1* encodes the major receptor of C3b, a protein

involved in complement activation, and could mediate complement-driven phagocytosis that fosters Aß clearance (for summary see ref. [7]). *PICALM* (aka *CALM*) encodes phosphatidylinositol-binding clathrin assembly protein which is involved in clathrin-mediated endocytosis,[21, 22] a process that could alter risk for AD through an effect on synaptic transmission or by altering endocytosis of the amyloid beta protein precursor. Thus all three new genes afford good opportunities for pursuit in biological experiments aimed at identifying novel approaches to the therapy of AD.

Acknowledgements

We thank contributors, including the Alzheimer's disease centers who collected samples used in this study, as well as subjects and their families, whose help and participation made this work possible. The principal investigator had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Author contributions

MMC directed this study and performed data analysis, OB took primary responsibility for drafting the manuscript assisted by MMC and KM, TAH performed the genotyping, LM, was responsible for DNA sample preparation, quality control and generated all DNA replica plates. JEC and VSP provided statistical expertise. D.W.D. is the pathologist who diagnosed and provided brain samples for the autopsy-confirmed

(AUT) series. N.R.G.-R. and R.C.P. are the neurologists who diagnosed and provided samples for the Mayo Clinic Jacksonville (JS) and Mayo Clinic Rochester (RS) series, respectively. KM participated in critical revisions of the manuscript, SGY is the lead investigator of this study.

Funding

Support for this research was provided by grants from the US National Institutes of Health, NIA (R01 AG18023, NRG-R, SGY); Mayo Alzheimer's Disease Research Center, (P50 AG16574, RCP, DWD, NRG-R, SGY); Mayo Alzheimer's Disease Patient Registry, (U01 AG06576, RCP); and US National Institute on Aging (AG25711, AG17216, AG03949, DWD). Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. This project was also generously supported by the Robert and Clarice Smith Postdoctoral Fellowship (MMC); Robert and Clarice Smith and Abigail Van Buren Alzheimer's Disease Research Program (RCP, DWD, NRG-R; SGY) and by the Palumbo Professorship in Alzheimer's Disease Research (SGY). KM is funded by the Alzheimer's Research Trust and the Big Lottery Fund, OB is partially funded by a travel grant from the Alzheimer's Research Trust. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Bertram L, Tanzi RE. The current status of Alzheimer's disease genetics: what do we tell the patients? *Pharmacol Res.* 2004;50(4):385-396.
2. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry.* 2006;63(2):168-174.
3. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science.* 1993;261(5123):921-923.
4. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology.* 1993;43(8):1467-1472.
5. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *Jama.* 1997;278(16):1349-1356.
6. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet.* 2009;41(10):1088-1093.
7. Lambert JC, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet.* 2009;41(10):1094-1099.
8. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group

under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology.* 1984;34(7):939-944.

9. Bertram L, Tanzi RE. Genome-wide association studies in Alzheimer's disease. *Hum Mol Genet.* 2009;18(R2):R137-145.

10. Carrasquillo MM, Zou F, Pankratz VS, et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat Genet.* 2009;41(2):192-198.

11. 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(1):17-23. Accessed on 11/30/09.

12. Rogaeva E, Meng Y, Lee JH, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet.* 2007;39(2):168-177.

13. Ghiso J, Matsubara E, Koudinov A, et al. The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is complexed to SP-40,40 (apolipoprotein J), an inhibitor of the complement membrane-attack complex. *Biochem J.* 1993;293 (Pt 1)(27-30.

14. Zlokovic BV, Martel CL, Mackic JB, et al. Brain uptake of circulating apolipoproteins J and E complexed to Alzheimer's amyloid beta. *Biochem Biophys Res Commun.* 1994;205(2):1431-1437.

15. Golabek A, Marques MA, Lalowski M, Wisniewski T. Amyloid beta binding proteins in vitro and in normal human cerebrospinal fluid. *Neurosci Lett.* 1995;191(1-2):79-82.

16. Zlokovic BV, Martel CL, Matsubara E, et al. Glycoprotein 330/megalin: probable role in receptor-mediated transport of apolipoprotein J alone and in a complex with Alzheimer disease amyloid beta at the blood-brain and blood-cerebrospinal fluid barriers. *Proc Natl Acad Sci U S A.* 1996;93(9):4229-4234.

17. Boggs LN, Fuson KS, Baez M, et al. Clusterin (Apo J) protects against in vitro amyloid-beta (1-40) neurotoxicity. *J Neurochem.* 1996;67(3):1324-1327.

18. DeMattos RB, O'Dell M A, Parsadanian M, et al. Clusterin promotes amyloid plaque formation and is critical for neuritic toxicity in a mouse model of Alzheimer's disease. *Proc Natl Acad Sci U S A.* 2002;99(16):10843-10848.

19. Lambert MP, Barlow AK, Chromy BA, et al. Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins. *Proc Natl Acad Sci U S A.* 1998;95(11):6448-6453.

20. Matsubara E, Soto C, Governale S, Frangione B, Ghiso J. Apolipoprotein J and Alzheimer's amyloid beta solubility. *Biochem J.* 1996;316 (Pt 2)(671-679.

21. Dreyling MH, Martinez-Climent JA, Zheng M, Mao J, Rowley JD, Bohlander SK. The t(10;11)(p13;q14) in the U937 cell line results in the fusion of the AF10 gene and CALM, encoding a new member of the AP-3 clathrin assembly protein family. *Proc Natl Acad Sci U S A.* 1996;93(10):4804-4809.

22. Tebar F, Bohlander SK, Sorkin A. Clathrin assembly lymphoid myeloid leukemia (CALM) protein: localization in endocytic-coated pits, interactions with clathrin, and the impact of overexpression on clathrin-mediated traffic. *Mol Biol Cell.* 1999;10(8):2687-2702.

Table 1. Association of variants in *CLU*, *CR1* and *PICALM* with LOAD in the original GWAS and in the follow-up series

Study	*N*		MAF			Association test	
	Cases	Controls	Cases	Controls	HWE	OR (95% CI)	p value[a]
***CLU* - rs11136000 - T (minor) allele**							
Harold GWAS	3,912	7,844	0.36	0.40	$6.1x10^{-1}$	0.84 (0.79-0.89)	$1.4x10^{-9}$
Lambert GWAS	2,016	5,266	0.35	0.39	$6.0x10^{-1}$	0.83 (0.77-0.90)	$1.5x10^{-6}$
Harold Follow-up	N/A	N/A	N/A	N/A	N/A	0.91 (0.83-0.99)	$1.7x10^{-2}$
Lambert Follow-up	3,775	3,154	0.35	0.38	$2.6x10^{-1}$	0.88 (0.81-0.95)	$8.8x10^{-4}$
Mayo[b]	1,819	2,565	0.37	0.41	$3.8x10^{-1}$	0.82 (0.75-0.91)	$8.6x10^{-5}$
JS	501	625	0.37	0.41	$6.0x10^{-1}$	0.82 (0.68-0.99)	$3.6x10^{-2}$
RS	314	1,635	0.40	0.41	$6.7x10^{-1}$	0.95 (0.78-1.15)	$5.9x10^{-1}$
AUT	315	99	0.36	0.36	$6.7x10^{-1}$	1.00 (0.69-1.43)	$9.8x10^{-1}$
NCRAD	689	206	0.35	0.41	$1.5x10^{-1}$	0.71 (0.54-0.92)	$1.0x10^{-2}$
Harold Combined						0.86 (0.82-0.90)	$8.5x10^{-10}$
Lambert Combined						0.86 (0.81-0.90)	$7.5x10^{-9}$
Harold/Lambert[c]							$3.1x10^{-20}$
Harold/Lambert/Mayo[c]							$2.7x10^{-26}$
***CR1*- rs3818361 - A (minor) allele**							
Harold GWAS	3,939	7,848	0.21	0.18	$8.3x10^{-2}$	1.17 (1.09-1.25)	$9.2x10^{-6}$
Lambert GWAS	2,018	5,324	0.22	0.18	$8.5x10^{-1}$	1.28 (1.17-1.40)	$8.5x10^{-8}$
Lambert Follow-up	3,717	3,094	0.22	0.20	$1.1x10^{-1}$	1.11 (1.02-1.22)	$1.6x10^{-2}$
Mayo[b]	1,829	2,576	0.22	0.20	$3.2x10^{-2}$	1.15 (1.03-1.29)	$1.4x10^{-2}$
JS	500	625	0.22	0.2	$1.8x10^{-1}$	1.19 (0.96-1.48)	$1.1x10^{-1}$
RS	313	1,641	0.19	0.19	$9.5x10^{-2}$	0.98 (0.77-1.24)	$8.6x10^{-1}$
AUT	316	102	0.22	0.21	$4.2x10^{-1}$	1.21 (0.78-1.87)	$4.1x10^{-1}$
NCRAD	700	208	0.23	0.19	$2.2x10^{-1}$	1.18 (0.86-1.61)	$3.1x10^{-1}$
Harold Combined						1.17 (1.09-1.25)	$9.2x10^{-6}$
Lambert Combined						1.19 (1.11-1.26)	$8.9x10^{-8}$
Harold/Lambert[c]							$1.3x10^{-14}$
Harold/Lambert/Mayo[c]							$1.8x10^{-16}$
***PICALM* - rs3851179 - T (minor) allele**							
Harold GWAS	3,941	7,848	0.33	0.37	$6.8x10^{-1}$	0.85 (0.80-0.90)	$1.9x10^{-8}$
Harold Follow-up	N/A	N/A	N/A	N/A	N/A	0.90 (0.82-0.99)	$1.4x10^{-2}$
Mayo[b]	1,816	2,552	0.33	0.37	$7.0x10^{-1}$	0.80 (0.73-0.89)	$1.3x10^{-5}$
JS	498	622	0.31	0.36	$6.0x10^{-1}$	0.76 (0.63-0.92)	$5.0x10^{-3}$
RS	313	1,622	0.33	0.38	$5.6x10^{-1}$	0.76 (0.62-0.94)	$1.0x10^{-2}$
AUT	308	100	0.35	0.40	$2.0x10^{-3}$	0.73 (0.52-1.03)	$7.6x10^{-2}$
NCRAD	697	208	0.33	0.36	$9.9x10^{-1}$	0.97 (0.73-1.27)	$8.1x10^{-1}$
Harold Combined						0.86 (0.82-0.90)	$1.3x10^{-9}$
Harold/Mayo[c]							$3.5x10^{-15}$

Abbreviations: MAF, minor allele frequency; HWE, p value for the test of Hardy-Weinberg equilibrium in controls; OR, odds ratio for the minor allele; 95% CI, 95% confidence interval.

[a]p values and ORs were calculated under an allelic dosage model using logistic regression adjusted by sex and *APOE ε4* status.

[b]The Mayo series reported here is independent of that which was incorporated in the GWAS reported by Harold *et al*.

[c]Indicates Fisher's combined p value.