Title
Novel late-onset Alzheimer disease loci variants associate with brain gene expression.

Journal
Neurology, 79(3)

ISSN
0028-3878

Authors
Allen, Mariet
Zou, Fanggeng
Chai, High Seng
et al.

Publication Date
2012-07-01

DOI
10.1212/wnl.0b013e3182605801

Peer reviewed

Novel late-onset Alzheimer disease loci variants associate with brain gene expression

Mariet Allen, PhD*
Fanggeng Zou, PhD*
High Seng Chai, PhD*
Curtis S. Younkin*
Julia Crook, PhD
V. Shane Pankratz, PhD
Minerva M. Carrasquillo, PhD
Christopher N. Rowley
Asha A. Nair
Sumit Middha
Sooraj Maharjan
Thuy Nguyen
Li Ma
Kimberly G. Malphrus
Ryan Palusak
Sarah Lincoln
Gina Bisceglio
Constantin Georgescu, PhD
Debra Schultz
Fariborz Rakhshan
Christopher P. Kolbert
Jin Jen, PhD
Jonathan L. Haines, PhD
Richard Mayeux, MD
Margaret A. Pericak-Vance, PhD
Lindsay A. Farrer, PhD
Gerard D. Schellenberg, PhD
Alzheimer's Disease Genetics Consortium
Ronald C. Petersen, MD, PhD
Neill R. Graff-Radford, MD
Dennis W. Dickson, MD
Steven G. Younkin, MD, PhD
Nilüfer Ertekin-Taner, MD, PhD

Correspondence & reprint requests to Dr. Ertekin-Taner: taner.nilufer@mayo.edu

Editorial, page 204

Supplemental data at www.neurology.org

Supplemental Data

ABSTRACT

Objective: Recent genome-wide association studies (GWAS) of late-onset Alzheimer disease (LOAD) identified 9 novel risk loci. Discovery of functional variants within genes at these loci is required to confirm their role in Alzheimer disease (AD). Single nucleotide polymorphisms that influence gene expression (eSNPs) constitute an important class of functional variants. We therefore investigated the influence of the novel LOAD risk loci on human brain gene expression.

Methods: We measured gene expression levels in the cerebellum and temporal cortex of autopsied AD subjects and those with other brain pathologies (~400 total subjects). To determine whether any of the novel LOAD risk variants are eSNPs, we tested their *cis*-association with expression of 6 nearby LOAD candidate genes detectable in human brain (*ABCA7, BIN1, CLU, MS4A4A, MS4A6A, PICALM*) and an additional 13 genes ±100 kb of these SNPs. To identify additional eSNPs that influence brain gene expression levels of the novel candidate LOAD genes, we identified SNPs ±100 kb of their location and tested for *cis*-associations.

Results: *CLU* rs11136000 ($p = 7.81 \times 10^{-4}$) and *MS4A4A* rs2304933/rs2304935 ($p = 1.48 \times 10^{-4}$–1.86×10^{-4}) significantly influence temporal cortex expression levels of these genes. The LOAD-protective *CLU* and risky *MS4A4A* locus alleles associate with higher brain levels of these genes. There are other *cis*-variants that significantly influence brain expression of *CLU* and *ABCA7* ($p = 4.01 \times 10^{-5}$–9.09×10^{-9}), some of which also associate with AD risk ($p = 2.64 \times 10^{-2}$–6.25×10^{-5}).

Conclusions: *CLU* and *MS4A4A* eSNPs may at least partly explain the LOAD risk association at these loci. *CLU* and *ABCA7* may harbor additional strong eSNPs. These results have implications in the search for functional variants at the novel LOAD risk loci. ***Neurology*® 2012;79:221–228**

GLOSSARY

ABC = ATP-binding cassette; **AD** = Alzheimer disease; **ADGC** = Alzheimer's Disease Genetics Consortium; **CER** = cerebellar tissue; **eSNP** = single nucleotide polymorphisms that influence gene expression; **GWAS** = genome-wide association study; **LOAD** = late-onset Alzheimer disease; **QC** = quality control; **RIN** = RNA Integrity Number; **TCX** = temporal cortex.

Recent genome-wide association studies (GWAS) of late-onset Alzheimer disease (LOAD) identified 9 novel loci, in addition to *APOE*.[1–6] Despite the success of disease GWAS, there remains a knowledge gap with this approach. First, loci identified by disease GWAS may harbor more than one candidate gene.[7] Second, the mechanism of action of the risk variants at the disease loci is not immediately obvious.[1,7] Third, variants identified in complex disease GWAS do not explain all their underlying genetic components.[8,9]

*These authors contributed equally to this work.

From the Department of Neuroscience (M.A., F.Z., C.S.Y., M.M.C., C.N.R., T.N., L.M., K.G.M., R.P., S.L., G.B., C.G., D.W.D., S.G.Y., N.E.-T.), Biostatistics Unit (J.C.), and Department of Neurology (N.R.G.-R., N.E.-T.), Mayo Clinic Florida, Jacksonville; Department of Biostatistics (H.S.C., V.S.P., A.A.N., S. Middha, S. Maharjan), Gene Expression Core, Advanced Genome Technology Center (D.S., F.R., C.P.K., J.J.), and Department of Neurology (R.C.P.), Mayo Clinic Minnesota, Rochester; Department of Molecular Physiology and Biophysics and Vanderbilt Center for Human Genetics Research (J.L.H.), Vanderbilt University, Nashville, TN; Gertrude H. Sergievsky Center, Department of Neurology, and Taub Institute on Alzheimer's Disease and the Aging Brain (R.M.), Columbia University, New York, NY; The John P. Hussman Institute for Human Genomics and Dr. John T. MacDonald Foundation Department of Human Genetics (M.A.P.-V.), University of Miami, Miami, FL; Departments of Biostatistics, Medicine (Genetics Program), Ophthalmology, Neurology, and Epidemiology (L.A.F.), Boston University, Boston, MA; and Perelman School of Medicine, Department of Pathology and Laboratory Medicine (G.D.S.), University of Pennsylvania, Philadelphia.

Alzheimer's Disease Genetics Consortium (ADGC) authors and their funding, acknowledgments, disclosure statements, and affiliations are listed on the *Neurology*® Web site at www.neurology.org.

Study funding: Funding information is provided at the end of the article.

Go to Neurology.org for full disclosures. Disclosures deemed relevant by the authors, if any, are provided at the end of this article.

Table 1 Subjects with temporal cortex and cerebellar expression measurements

	Temporal cortex		Cerebellum	
	AD	Non-AD	AD	Non-AD
No.	202	197	197	177
Female, n (%)	108 (53)	78 (40)	101 (51)	63 (36)
Mean age, y, ±SD	73.6 ± 5.5	71.6 ± 5.6	73.6 ± 5.6	71.7 ± 5.5
ApoE4 dose 0, n (%)	79 (39)	146 (74)	71 (36)	130 (73)
ApoE4 dose 1, n (%)	96 (48)	46 (23)	98 (50)	43 (24)
ApoE4 dose 2, n (%)	27 (13)	3 (2)	28 (14)	2 (1)
ApoE4 dose unknown, n (%)	0 (0)	2 (1)	0 (0)	2 (1)
Mean RIN ± SD	6.3 ± 0.9	6.9 ± 1.0	7.2 ± 1	7.2 ± 1

Abbreviations: AD = Alzheimer disease; RIN = RNA Integrity Number.

An alternative approach is using biologically relevant, genetically driven, quantitative phenotypes, i.e., endophenotypes. Gene expression levels are a special group of endophenotypes with a substantial genetic component.[7,10–17] Identification of variants that influence both gene expression and human disease can discover the actual risk genes at the GWAS loci and their potential mechanism of action.[7,10,11,13,14,16]

We postulate that LOAD has multiple risk variants, some of which influence gene expression (eSNPs). In this study, we assessed the novel LOAD GWAS loci, first to determine if any of the "top LOAD risk SNPs" influence brain expression of nearby genes in-*cis*. We arbitrarily defined *cis*-associations as those between a transcript and a SNP, which is within the gene encoding that transcript or within 100 kb of its 5′ start or 3′ end sites. Second, we aimed to determine if the brain expression levels of any of the "top LOAD candidate genes" at these loci are influenced by any other eSNPs in-*cis* (*cis*-eSNPs). In this study, we utilize the term eSNPs to define variants associating with expression levels and not to indicate direct evidence of functionality, as the eSNPs are likely markers of other functional, regulatory variants. To determine whether any of the significant eSNPs that are identified in this second aim also influence AD risk, we utilized meta-analyses results from the LOAD risk GWAS of the Alzheimer's Disease Genetics Consortium (ADGC).[6] Our findings have implications about the potential mechanism of action for some of the top LOAD risk SNPs and for functional regulation of some of the top LOAD candidate genes.

METHODS

Subjects. We measured gene expression levels from temporal cortex (TCX) of 399 subjects and from cerebellar tissue (CER) of 374 subjects (table 1). A total of 340 subjects had both TCX and CER measurements. All subjects were participants in the published Mayo LOAD GWAS[18] as part of the autopsy-based series (AUT_60–80). All subjects had neuropathologic evaluation by D.W.D. All subjects with AD (n = 202 for TCX and 197 for CER) had definite diagnosis of AD according to the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association criteria[19] and had Braak scores of ≥4.0. All control subjects (non-AD, n = 197 for TCX and 177 for CER) had Braak scores of ≤2.5, but many had brain pathology unrelated to AD (table e-1 on the *Neurology*® Web site at www.neurology.org).

RNA extraction and gene expression measurements. Total RNA was extracted from frozen TCX or CER samples using the Ambion RNAqueous kit according to the manufacturer's instructions. The quantity and quality of the RNA samples were determined by the Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Nano Chip.

Transcript levels were measured using the Whole Genome DASL assay (Illumina, San Diego, CA). The RNA samples were randomized across the chips and plates using a stratified approach to ensure balance with respect to diagnosis, age, gender, RNA Integrity Numbers (RINs), and *APOE* genotype. Replicate samples were utilized for quality control (QC). Raw probe-level mRNA expression data of the 399 TCX and 374 CER samples were exported from GenomeStudio software (Illumina Inc.) for preprocessing with background correction, variance stabilizing transformation, quantile normalization, and probe filtering using the lumi package of BioConductor[24,25] (appendix e-2: Methods).

Standard protocol approvals, registrations, and patient consents. This study was approved by the appropriate institutional review board and appropriate informed consent was obtained from all participants.

Genotype data. All genotypes used in the SNP/transcript level association studies were extracted from the Mayo LOAD GWAS, since all autopsied subjects with TCX or CER gene expression measurements were part of this study.[18] The LOAD GWAS genotypes were generated using Illumina's Human Hap300-Duo Genotyping BeadChips analyzed with an Illumina BeadLab Station (Illumina) at the Mayo Clinic Genotyping Shared Resource according to the manufacturer's protocols. The LOAD GWAS QC methods were previously published[10,18] (appendix e-2: Methods).

Statistical methods. ***SNP/transcript levels association.*** To test the influence of the top LOAD risk SNPs on brain gene expression in-*cis*, we extracted from the Mayo LOAD GWAS the genotypes for these SNPs or their proxies. The Mayo LOAD GWAS had genotypes for the following 5 top LOAD risk SNPs: *BIN1*_rs744373, *CLU*_rs1136000, *EPHA1*_rs11767557, *MS4A6A*_rs610932, and *PICALM*_rs3851179; and the proxies for *ABCA7*_rs3764650, *MS4A4A*_rs4938933, *MS4A6A*_rs610932, *MS4A4E*_rs670139, and *PICALM*_rs561655 (table e-2). The top LOAD risk SNPs or their proxies were tested for *cis*-association with the transcript

levels of those genes that resided within ±100 kb of their genomic location according to NCBI Build 36. Of the top LOAD candidate genes nearby these SNPs, the following had detectable brain expression levels in both TCX and CER: *ABCA7*, *BIN1*, *CLU*, *MS4A4A*, *MS4A6A*, and *PICALM*. Thirteen additional nearby genes were also tested for SNP/transcript levels associations. A total of 59 SNP/transcript associations were tested (tables e-2 and e-3).

To test whether the brain expression levels of any of the top LOAD candidate genes are influenced by other eSNPs in-*cis* (*cis*-eSNPs), we identified all SNPs in the Mayo LOAD GWAS within ±100 kb flanking region of the gene. *ABCA7*, *BIN1*, *CLU*, *MS4A4A*, *MS4A6A*, *PICALM*, and *CD2AP* were tested for additional eSNPs. A total of 369 SNP/transcript associations were tested (table e-4 and e-5). Study-wide Bonferroni corrections were done for the total number of SNP/transcript associations tested.

Linear regression analysis to test for SNP/transcript associations were done in PLINK.[20] Preprocessed probe transcript levels were utilized as endophenotypes. Each probe was assessed separately, even though 1 gene may have multiple probes. An additive model was assumed, with the minor allele dosage (0, 1, 2) as the independent variable, and *APOE* ϵ4 dosage (0, 1, 2), age at death, gender, PCR plate, RIN, and adjusted RIN^2 (defined as [RIN − RINmean]^2) as covariates. The CER and TCX results were analyzed separately. The AD and non-AD subjects were analyzed both separately and jointly. The joint analyses including diagnosis as an additional covariate (AD = 1, non-AD = 0) were considered to be the main analyses. The separate AD and non-AD only analyses are also reported (tables e-1 through e-6).

ADGC meta-analyses. The top LOAD risk SNPs that were assessed for their influence on brain gene expression have AD risk association results published in the literature. To determine the AD risk association of the remaining SNPs assessed in this study (i.e., proxies for the top LOAD risk SNPs and the other *cis*-eSNPs), we obtained meta-analyses results from the ADGC.[6] The cohorts that are assessed by ADGC, as well as the methodologic details of the meta-analyses, are described in detail in a recent publication.[6] Briefly, the meta-analyses of the ADGC dataset results reported here (tables 2, 3 and e-2 through e-5) are generated from the combined analyses of stage 1 and stage 2 cohorts (appendix e-2: Methods), with detailed descriptions provided elsewhere.[6,21] Stage 1 cohorts are comprised of 8,309 LOAD cases and 7,366 cognitively normal elder controls. Stage 2 has 3,531 LOAD vs 3,565 control subjects.

Each cohort was tested for AD risk association using a logistic regression approach, assuming an additive model and adjusting for age, sex, *APOE* ϵ4 dosage, and principal components from EIGENSTRAT.[22] The meta-analyses results were generated using the inverse variance method implemented in the software package METAL.[23]

RESULTS *cis*-Association of the novel top LOAD risk SNPs with brain expression levels of genes. We evaluated 18 top LOAD risk SNPs or their proxies from 6 novel LOAD loci for associations with TCX and CER expression levels of 6 top LOAD candidate genes and 13 others within ±100 kb of these SNPs. All 19 genes had detectable expression both in TCX and CER of autopsied subjects. Given the assessment of 59 SNP/transcript associations (table e-2), $p < 8.47 \times 10^{-4}$ is required to achieve significance after Bonferroni correction. *CLU*_rs11136000 and 2

Table 2 Significant expression level associations for the top AD locus SNPs[a]

CHR	eSNP	Tested allele	DASL expression probe	Symbol[b]	TX all		CER all		AD-locus association[c]					LD (eSNP and AD SNP)		AD-eSNP association[e]	
					p	*β*	*p*	*β*	AD locus	AD SNP[d]	OR	*p*	Reference	R^2	D	OR (95% CI)	*p*
11	rs2304933	A	ILMN_2370336	MS4A4A	1.48E-04	0.22	3.65E-02	0.09	MS4A4E	rs670139	1.09	1.40E-09	Hollingworth et al.[5]	0.621	0.907	1.06 (1.01-1.11)	1.13E-02
11	rs2304935	G	ILMN_2370336	MS4A4A	1.86E-04	0.22	3.65E-02	0.09	MS4A4E	rs670139	1.09	1.40E-09	Hollingworth et al.[5]	0.621	0.907	1.06 (1.01-1.11)	1.11E-02
8	rs11136000	A	ILMN_1667058	CLU	7.81E-04	0.17	NS	0.03	CLU	rs11136000	0.84	1.40E-09	Harold/Lambert et al.[2,3]	NA	NA	0.89 (0.85-0.93)	5.226E-07

Abbreviations: AD = Alzheimer disease; CER = cerebellar tissue; CER all = eSNP/expression associations in the cerebellar tissue of all subjects; CI = confidence interval; eSNP = single nucleotide polymorphisms tested for expression level associations; LD = linkage disequilibrium (if eSNP and AD SNP are different, then their LD is depicted with R^2 and D'); NA = not applicable; NS = not significant; OR = odds ratio; TCX = temporal cortex; TX all = eSNP/expression associations in the TCX of all subjects.

[a] The results that have study-wide significance for TCX, CER, or both after Bonferroni correction for 59 SNP/transcript association tests are depicted.

[b] Gene tested for its brain expression level association with the eSNP.

[c] Published AD risk–locus association results.

[d] Published AD locus SNP with strongest AD risk association. This SNP may be different from the eSNP.

[e] Association of the eSNP with AD risk in the meta-analysis of the ADGC dataset.

Table 3 Significant associations for the *cis*-eSNPs and brain expression levels of the top AD loci genes[a]

CHR	eSNP	Tested allele	DASL expression probe	Symbol[b]	TX all		CER all		AD-eSNP association[c]	
					p	β	*p*	β	OR (95% CI)	*p*
8	rs894019	A	ILMN_1667058	CLU	9.00E-09	0.28	NS	0.01	1.00 (0.95–1.05)	NS
8	rs569214	A	ILMN_1667058	CLU	1.98E-06	0.24	NS	−0.02	0.95 (0.91–0.99)	2.64E-02
8	rs542876	A	ILMN_1667058	CLU	2.13E-06	0.21	NS	−0.01	NA	NA
8	rs473024	G	ILMN_1667058	CLU	1.04E-05	0.24	NS	0.06	1.04 (0.99–1.10)	NS
8	rs2582369	A	ILMN_1667058	CLU	2.02E-05	0.23	NS	0.07	1.04 (0.98–1.09)	NS
19	rs7247087	A	ILMN_1743205	ABCA7	3.53E-05	0.14	1.03E-07	0.18	0.97 (0.91–1.03)	NS
8	rs570197	G	ILMN_1667058	CLU	4.01E-05	0.22	NS	0.06	1.03 (0.98–1.09)	NS
19	rs757232	A	ILMN_1743205	ABCA7	1.61E-03	0.10	2.62E-06	0.15	1.11 (1.05–1.17)	6.25E-05
19	rs2072102	A	ILMN_1743205	ABCA7	1.12E-03	0.11	3.57E-06	0.15	1.11 (1.05–1.17)	6.74E-05

Abbreviations: AD = Alzheimer disease; CER = cerebellar tissue; CER all = eSNP/expression associations in the CER of all subjects; CI = confidence interval; eSNP = single nucleotide polymorphisms tested for expression level associations; NA = not applicable; NS = not significant; OR = odds ratio; SNP = single nucleotide polymorphism; TCX = temporal cortex; TX all = eSNP/expression associations in the TCX of all subjects.

[a] The results that have study-wide significance for TCX, CER, or both after Bonferroni correction for 369 SNP/transcript association tests are depicted.

[b] Gene tested for its brain expression level association with the eSNP.

[c] Association of the eSNP with AD risk in the meta-analysis of the Alzheimer's Disease Genetics Consortium dataset.

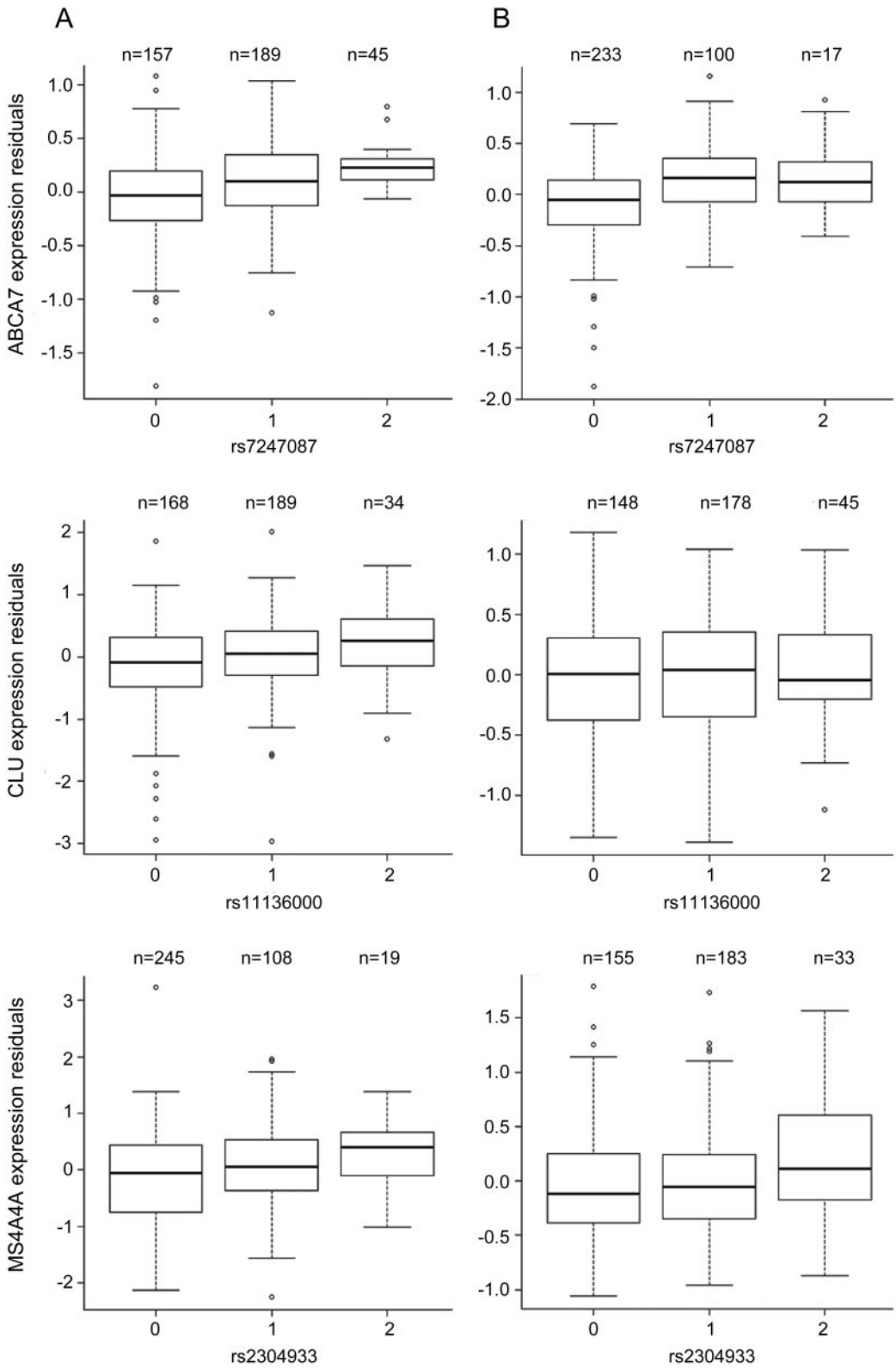
SNPs at the *MS4A* locus (rs2304933 and rs2304935) have significant association with TCX expression levels of *CLU* ($p = 7.81 \times 10^{-4}$) and *MS4A4A* ($p = 1.48 \times 10^{-4}$ and 1.86×10^{-4}), respectively (table 2 and figures e-1 and e-2). The *MS4A4A* eSNPs also have nominally significant associations in CER ($p = 3.65 \times 10^{-2}$), but *CLU*_rs11136000/transcript association is not significant in CER. The significant *MS4A4A* SNP/transcript associations appear to be driven by both AD and non-AD subjects, however *CLU*_rs11136000/transcript association achieves significance only in the non-ADs (tables e-2 and e-3).

*CLU*_rs11136000 minor "A" allele associates with both higher TCX expression levels of this gene and lower AD risk (tables 2, e-2, e-3, and figure 1A). *CLU*_rs11136000 reached genome-wide significance in published LOAD GWAS[2,3] and also has significant association with lower AD risk in the meta-analyses of the ADGC dataset (OR = 0.89, $p = 5.23 \times 10^{-7}$). The minor alleles of *MS4A4A* SNPs rs2304933 and rs2304935 associate with higher brain expression levels of this and increased AD risk that is nominally significant in the meta-analysis of the ADGC dataset (OR = 1.06, $p = 1.11 \times 10^{-2}$ to 1.13×10^{-2}) (table 2, table e-3, figure 1, A and B). The *MS4A4A* transcript-associating SNPs rs2304933 and rs2304935 are proxies for rs670139, which is intergenic between *MS4A4E* and *MS4A6A* and which showed genome-wide significant association with increased AD risk.[5]

***cis*-Association of SNPs with brain expression levels of novel top LOAD candidate genes.** We evaluated expression levels of 7 novel top LOAD candidate genes, at 6 of the recently identified novel LOAD loci, detectable in TCX and CER of autopsied subjects, for their association with *cis*-eSNPs residing within ±100 kb of their flanking region. Given the 369 SNP/transcript associations tested, $p < 1.36 \times 10^{-4}$ is needed to achieve Bonferroni-corrected study-wide significance. There were 9 significant SNP/transcript associations; 6 for *CLU* expression in TCX ($p = 4.01 \times 10^{-5}$ to 9.00×10^{-9}), and 3 for *ABCA7* in both TCX and CER ($p = 3.53 \times 10^{-5}$ to 1.03×10^{-7}) (tables 3, e-4, and e-5 and figures e-1 and e-3). *ABCA7*_rs7247087/transcript associations were significant in both brain regions, and the other 2 *ABCA7* SNPs (rs757232 and rs2072102) that were significant in CER had nominal significance in TCX. The direction and magnitude of the effect on *ABCA7* levels were similar in the 2 brain regions for all 3 SNPs. None of the 5 *CLU* SNPs with significant TCX transcript associations showed nominal significance in CER. All significant *CLU* and *ABCA7* SNP/transcript associations showed effects of similar direction and magnitude in both AD and non-AD brains.

*CLU*_ rs569214 which is associated with higher levels of this gene in TCX has nominally significant association with lower AD risk in the meta-analysis of the ADGC dataset (table 3, table e-5, figure 1A). *ABCA7*_ rs757232 and rs2072102 show significant association with higher *ABCA7* levels in 2 brain regions as well as significantly higher risk of AD (OR = 1.11, $p = 6.25 \times 10^{-5}$ to 6.74×10^{-5}) in the meta-analysis of the ADGC dataset (table 3, table e-5, figure 1, A and B). None of the other significant

Figure 1 Box plots of the residuals for the preprocessed expression values for *ABCA7*, *CLU*, and *MS4A4A* genes in the (A) temporal cortex and (B) cerebellum



Residuals were obtained following adjustment for age at death, gender, *APOE* ϵ4 dose, PCR plate, RNA Integrity Number (RIN), adjusted RIN^2, and diagnosis in all subjects included in the analysis, for each genotype (0, 1, or 2 alleles) of the targeted SNP (rs#) in a linear regression model. Median values are represented by a thick, black, horizontal line within the box, while the box represents the upper and lower quartiles. The whiskers represent the maximum and minimum values (excluding outliers) defined as 1.5 times the interquartile range. Outliers are represented as circles. The subject counts for each genotype are indicated (N =) above each box.

eSNPs showed any association with AD risk in this study.

DISCUSSION The novel disease loci identified in the recent large disease GWAS provide the opportunity to uncover the pathophysiology of complex diseases, such as LOAD, which may be instrumental in the discovery of novel drug targets and development of early diagnostic tools. Nonetheless, significant progress awaits the identification of the true disease genes at these loci, their functional risk alleles and mechanisms of action. In this study, we used the brain gene expression endophenotype and disease GWAS results in a combined fashion to test whether any of the top LOAD risk SNPs (referred henceforth as LOAD SNPs for short) are eSNPs and whether the expression of any of the top LOAD candidate genes is influenced by any other *cis*-eSNPs that are within ±100 kb of their genomic location.

Our approach revealed several important findings. First, the LOAD SNPs at the *CLU* and *MS4A* loci also influence brain gene expression of *CLU* and *MS4A4A* genes, respectively. Second, there are additional variants within the arbitrary 100 kb *cis*-region that also influence brain expression of *CLU* and *ABCA7*. Third, these additional, strong *CLU* and *ABCA7 cis*-eSNPs do not appear to be in strong LD with the LOAD SNPs at their respective loci.

The strongest LOAD SNP at the *ABCA7* locus, rs3764650, was assessed by its proxy rs375229 in our study and not found to have a significant association with brain *ABCA7* levels. Conversely, 2 of the strong *ABCA7 cis*-eSNPs (rs757232, rs2072102) also have significant AD risk association ($p = 6.25 \times 10^{-5}$ to 6.74×10^{-5}). For the *CLU* locus, the top LOAD SNP, rs11136000, also associates with *CLU* levels in the temporal cortex. Of the additional strong *CLU cis*-eSNPs identified in this study, rs569214 associates with AD risk with modest nominal significance ($p = 2.64 \times 10^{-2}$). That the *ABCA7* and *CLU* loci LOAD SNPs and additional *cis*-eSNPs, which also associate with AD risk, are not in strong LD with each other suggests that there may be multiple, independent variants at these loci that influence AD risk. Furthermore, at least some of this AD risk is conferred by regulatory variants which affect brain gene expression levels.

The direction of the *CLU* SNP/transcript and SNP/AD risk associations are biologically congruent. The top LOAD risk SNP rs11136000 and the other *CLU* transcript-associating SNPs have alleles which associate with increased brain levels of this gene and also reduce AD risk. *CLU* encodes for clusterin, which is thought to promote neuroprotection in AD via multifaceted functions including Aβ clearance, prevention of excessive inflammation, inhibition of apoptosis, and clearance of neuronal debris.[24] Thus, genetic variants that influence higher brain *CLU* levels could conceivably lower AD risk.

MS4A locus SNPs associate with higher levels of *MS4A4A* and increased AD risk. The functions of the *MS4A* locus genes, including *MS4A4A*, are yet to be established, though they are thought to be members of a family of transmembrane proteins which may be parts of oligomeric cell surface complexes involved in signal transduction.[25] Given that the *MS4A* locus harbors 6 genes in an LD block (*MS4A3*, *MS4A2*, *MS4A6A*, *MS4A4E*, *MS4A4A*, *MS4A6E*), it is not possible to distinguish the actual LOAD risk genes in this region based solely on the disease GWAS findings. Our findings suggest that the *MS4A4A* gene might be the strongest AD candidate at this locus, though they need to be interpreted with caution, since brain expression levels of only *MS4A4A* and *MS4A6A* could be tested. Thus, there may be other *MS4A* genes that are influenced by SNPs at this locus. Indeed, we found nominally significant SNP/transcript associations with *MS4A6A* and the downstream *MS4A7*. Others have previously shown nominally significant rs610932/*MS4A6A* transcript associations in brains of 143 neurologically normal, European subjects,[5] though we were not able to confirm this finding in our larger study.

The *ABCA7* transcript-associating SNPs both increase its brain levels and AD risk. ABCA7 is member of the ATP-binding cassette (ABC) family of proteins implicated in lipid metabolism,[26] shown in vitro to regulate cholesterol efflux, inhibit amyloid precursor protein processing,[27] and also to play a role in the phagocytosis of apoptotic cells.[28] ABCA7 has a splice variant with a distinct expression profile and which lacks the lipid metabolism functionality.[29] Understanding the implications of *ABCA7* expressional regulation in AD risk requires further information about the functions of this protein, and clarification about the brain *ABCA7* splice variant that is influenced by the *ABCA7* SNPs.

Our results also have implications regarding fine mapping studies aimed at functional variant discovery at the novel LOAD risk loci and suggest that screening only for coding polymorphisms may miss important functional variation. The strongest eSNP associations are in the 5′ or 3′ regions of the transcripts and sometimes in different genes. Bioinformatics focused on transcriptional regulation sites in noncoding regions within and surrounding the *CLU*, *MS4A*, and *ABCA7* genes, followed by targeted sequencing and in vitro functional efforts, may be fruitful in functional variant discovery at these loci.

It is clear that not all the functional disease variants will be eSNPs. We also note that the strongest

LOAD risk SNP was not the strongest eSNP and vice versa in this study. This could be due to differences in the populations tested for the LOAD risk and expression associations, sample sizes, LD patterns, or a combination of these factors. Nonetheless, it is also possible that the LOAD risk SNPs identified in the disease GWAS may be marking multiple different types of functional variants, i.e., both eSNPs and missense coding variants. Our results suggest that for 3 of the novel LOAD GWAS loci, eSNPs account for at least part of the disease risk. Not all novel LOAD candidate genes were detectable in both CER and TCX of our samples with the DASL microarrays. It will be important to investigate SNP associations for brain expression levels of the remaining LOAD candidate genes using alternative approaches. It is also necessary to characterize in detail the influence of the eSNPs on brain expression levels of any splice variants.

The strengths of our study include combined use of the gene expression endophenotype and disease GWAS, use of a relatively large sample size of both AD and non-AD brain tissue from 2 different regions, and investigation of both the top LOAD risk SNPs as well as other *cis*-eSNPs for their effects on brain expression of 7 top LOAD candidate genes. Our study demonstrates the power and utility of the gene expression endophenotypes; identifies brain expression changes in *CLU*, *MS4A4A*, and *ABCA7* as one potential mechanism of action at these novel LOAD risk loci; and provides direction about functional variant discovery at these loci.

ADGC AUTHORS

Liana G. Apostolova, MD, Steven E. Arnold, MD, Clinton T. Baldwin, PhD, Robert Barber, PhD, Michael M. Barmada, PhD, Thomas Beach, MD, PhD, Gary W. Beecham, PhD, Duane Beekly, BS, David A. Bennett, MD, Eileen H. Bigio, MD, Thomas D. Bird, MD, Deborah Blacker, MD, Bradley F. Boeve, MD, James D. Bowen, MD, Adam Boxer, MD, PhD, James R. Burke, MD, PhD, Jacqueline Buros, BS, Joseph D. Buxbaum, PhD, Nigel J. Cairns, PhD, FRCPath, Laura B. Cantwell, MPH, Chuanhai Cao, PhD, Chris S. Carlson, PhD, Regina M. Carney, MD, Steven L. Carroll, MD, PhD, Helena C. Chui, MD, David G. Clark, MD, Jason Corneveaux, BS, Carl W. Cotman, PhD, Paul K. Crane, MD, MPH, Carlos Cruchaga, PhD, Jeffrey L. Cummings, MD, Philip L. De Jager, MD, PhD, Charles DeCarli, MD, Steven T. DeKosky, MD, F. Yesim Demirci, MD, Ramon Diaz-Arrastia, MD, PhD, Malcolm Dick, PhD, Beth A. Dombroski, PhD, Ranjan Duara, MD, William G. Ellis, MD, Denis Evans, MD, Kelley M. Faber, MS, Kenneth B. Fallon, MD, Martin R. Farlow, MD, Steven Ferris, PhD, Tatiana M. Foroud, PhD, Matthew P. Frosch, MD, PhD, Douglas R. Galasko, MD, Paul J. Gallins, MS, Mary Ganguli, MD, Marla Gearing, PhD, Daniel H. Geschwind, MD, PhD, Bernardino Ghetti, MD, John R. Gilbert, PhD, Sid Gilman, MD, FRCP, Bruno Giordani, PhD, Jonathan D. Glass, MD, Alison M. Goate, D.Phil, Robert C. Green, MD, John H. Growdon, MD, Hakon Hakonarson, MD, PhD, Ronald L. Hamilton, MD, John Hardy, PhD, Lindy E. Harrell, MD, PhD, Elizabeth Head, PhD, Lawrence S. Honig, MD, PhD, Matthew J. Huentelman, PhD, Christine M. Hulette, MD, Bradley T. Hyman, MD, PhD, Gail P. Jarvik, MD, PhD, Gregory A. Jicha, MD, PhD, Lee-Way Jin, MD, PhD, Nancy Johnson, PhD, Gyungah Jun, PhD, M. Ilyas Kamboh, PhD, Jason Karlawish, MD, Anna Karydas, BA, John S.K. Kauwe, PhD, Jeffrey A. Kaye, MD, Ronald Kim, MD, Edward H. Koo, MD, Neil W. Kowall, MD, Patricia Kramer, PhD, Walter A. Kukull, PhD, James J. Lah, MD, PhD, Eric B. Larson, MD, MPH, Allan I. Levey, MD, PhD, Andrew P. Lieberman, MD, PhD, Oscar L. Lopez, MD, Kathryn L. Lunetta, PhD, Wendy J. Mack, PhD, Daniel C. Marson, JD, PhD, Eden R. Martin, PhD, Frank Martiniuk, PhD, Deborah C. Mash, PhD, Eliezer Masliah, MD, Wayne C. McCormick, MD, MPH, Susan M. McCurry, PhD, Andrew N. McDavid, BA, Ann C. McKee, MD, Marsel Mesulam, MD, Bruce L. Miller, MD, Carol A. Miller, MD, Joshua W. Miller, PhD, Thomas J. Montine, MD, PhD, John C. Morris, MD, Amanda J. Myers, PhD, Adam C. Naj, PhD, Petra Nowotny, PhD, Joseph E. Parisi, MD, Daniel P. Perl, MD, Elaine Peskind, MD, Wayne W. Poon, PhD, Huntington Potter, PhD, Joseph F. Quinn, MD, Ashok Raj, MD, Ruchita A. Rajbhandary, MPH, Murray Raskind, MD, Eric M. Reiman, MD, Barry Reisberg, MD, Christiane Reitz, MD, PhD, John M. Ringman, MD, Erik D. Roberson, MD, PhD, Ekaterina Rogaeva, PhD, Roger N. Rosenberg, MD, Mary Sano, PhD, Andrew J. Saykin, PsyD, Julie A. Schneider, MD, Lon S. Schneider, MD, William Seeley, MD, Michael L. Shelanski, MD, PhD, Michael A. Slifer, MD, PhD, Charles D. Smith, MD, Joshua A. Sonnen, MD, Salvatore Spina, MD, Peter St George-Hyslop, MD, FRCP, Robert A. Stern, PhD, Rudolph E. Tanzi, PhD, John Q. Trojanowski, MD, PhD, Juan C. Troncoso, MD, Debby W. Tsuang, MD, Vivianna M. Van Deerlin, MD, PhD, Badri Narayan Vardarajan, MS, Harry V. Vinters, MD, Jean Paul Vonsattel, MD, Li-San Wang, PhD, Sandra Weintraub, PhD, Kathleen A. Welsh-Bohmer, PhD, Jennifer Williamson, MS, Randall L. Woltjer, MD, PhD.

AUTHOR CONTRIBUTIONS

M.A.: drafting/revising the manuscript for content, study concept, analysis or interpretation of data. F.Z., H.S.C., C.S.Y., J.C., V.S.P., M.M.C.: drafting/revising the manuscript for content, analysis or interpretation of data. C.N.R., A.A.N., S. Middha, S. Maharjan, T.N., L.M., K.G.M., R.P., S.L., G.B., C.G., D.S., F.R.: analysis or interpretation of data. C.P.K., J.J., J.L.H., R.M., M.A.P.V., L.A.F., G.D.S.: drafting/revising the manuscript for content, analysis or interpretation of data. R.C.P., N.R.G.-R., D.W.D.: drafting/revising the manuscript for content. S.G.Y., N.E.T.: drafting/revising the manuscript for content, study concept, analysis or interpretation of data.

ACKNOWLEDGMENT

The authors thank the study participants and their families who made this work possible and members of the Mayo Clinic Jacksonville Memory Disorders Clinic and Mayo Clinic AD Research Center for their help in the collection of the samples.

STUDY FUNDING

Supported by National Institutes of Health grants: National Institute on Aging (R01 032990 to N.E.-T. and R01 AG018023 to N.R.G.-R. and S.G.Y.); Mayo Alzheimer's Disease Research Center (P50 AG016574 to R.C.P., D.W.D., N.R.G.-R., S.G.Y., and N.E.-T.); Mayo Alzheimer's Disease Patient Registry (U01 AG006576 to R.C.P.); National Institute on Aging (AG025711, AG017216, AG003949 to D.W.D.). Also supported by the Robert and Clarice Smith and Abigail Van Buren Alzheimer's Disease Research Program (to R.C.P., D.W.D., N.R.G.-R., and S.G.Y.), and by the Palumbo Professorship in Alzheimer's Disease Research (to S.G.Y.). N.E.T. is the recipient of National Institutes of Health (KL2 RR024151) and Siragusa Foundation grants.

DISCLOSURE

N. Graff-Radford has served as a consultant to Codman and received grant support from Elan Pharmaceutical Research, Pfizer Pharmaceuticals, Medivation, and Forrest. R.C. Petersen has been a consultant to GE Healthcare and Elan Pharmaceuticals, has served on a data safety monitoring board in a clinical trial sponsored by Elan Pharmaceuticals, and a safety monitoring board for Wyeth Pharmaceuticals. The remaining authors report no disclosures relevant to the manuscript. **Go to Neurology.org for full disclosures.**

Received August 22, 2011. Accepted in final form November 28, 2011.

REFERENCES

1. Ertekin-Taner N. Genetics of Alzheimer disease in the pre- and post-GWAS era. Alzheimers Res Ther 2010;2:3.
2. 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;41:1088–1093.
3. 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;41:1094–1099.
4. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 2010;303:1832–1840.
5. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 2011;43:429–435.
6. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 2011;43:436–441.
7. Ertekin-Taner N. Gene expression endophenotypes: a novel approach for gene discovery in Alzheimer's disease. Mol Neurodegener 2011;6:31.
8. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science 2008;322:881–888.
9. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–753.
10. Zou F, Carrasquillo MM, Pankratz VS, et al. Gene expression levels as endophenotypes in genome-wide association studies of Alzheimer disease. Neurology 2010;74:480–486.
11. Carrasquillo MM, Belbin O, Zou F, et al. Concordant association of insulin degrading enzyme gene (IDE) variants with IDE mRNA, Abeta, and Alzheimer's disease. PLoS One 2010;5:e8764.
12. Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. Nat Genet 2007;39:1217–1224.
13. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. Nat Genet 2007;39:1202–1207.
14. Goring HH, Curran JE, Johnson MP, et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat Genet 2007;39:1208–1216.
15. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. Nature 2008;452:423–428.
16. Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol 2008;6:e107.
17. Myers AJ, Gibbs JR, Webster JA, et al. A survey of genetic human cortical gene expression. Nat Genet 2007;39:1494–1499.
18. 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;41:192–198.
19. 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;34:939–944.
20. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
21. Jun G, Naj AC, Beecham GW, et al. Meta-analysis confirms CR1, CLU, and PICALM as Alzheimer disease risk loci and reveals interactions with APOE genotypes. Arch Neurol 2010;67:1473–1484. Erratum in Arch Neurol 2011;68:159.
22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–909.
23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191.
24. Nuutinen T, Suuronen T, Kauppinen A, Salminen A. Clusterin: a forgotten player in Alzheimer's disease. Brain Res Rev 2009;61:89–104.
25. Liang Y, Buckley TR, Tu L, Langdon SD, Tedder TF. Structural organization of the human MS4A gene cluster on chromosome 11q12. Immunogenetics 2001;53:357–368.
26. Abe-Dohmae S, Ueda K, Yokoyama S. ABCA7, a molecule with unknown function. FEBS Lett 2006;580:1178–1182.
27. Chan SL, Kim WS, Kwok JB, et al. ATP-binding cassette transporter A7 regulates processing of amyloid precursor protein in vitro. J Neurochem 2008;106:793–804.
28. Jehle AW, Gardai SJ, Li S, et al. ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. J Cell Biol 2006;174:547–556.
29. Ikeda Y, Abe-Dohmae S, Munehira Y, et al. Posttranscriptional regulation of human ABCA7 and its function for the apoA-I-dependent lipid release. Biochem Biophys Res Commun 2003;311:313–318.