Title

Gene-based aggregate SNP associations between candidate AD genes and cognitive decline

Journal

GeroScience, 38(2)

ISSN

2509-2715

Authors

Nettiksimmons, Jasmine
Tranah, Gregory
Evans, Daniel S
et al.

Publication Date

2016-04-01

DOI

10.1007/s11357-016-9885-2

Peer reviewed

AGE (2016) 38: 41
DOI 10.1007/s11357-016-9885-2

Gene-based aggregate SNP associations between candidate AD genes and cognitive decline

Jasmine Nettiksimmons · Gregory Tranah · Daniel S. Evans · Jennifer S. Yokoyama · Kristine Yaffe

Received: 9 July 2015 / Accepted: 28 January 2016 / Published online: 22 March 2016

Abstract Single nucleotide polymorphisms (SNPs) in and near *ABCA7*, *BIN1*, *CASS4*, *CD2AP*, *CD33*, *CELF1*, *CLU*, complement receptor 1 (*CR1*), *EPHA1*, *EXOC3L2*, *FERMT2*, *HLA* cluster *(DRB5-DQA)*, *INPP5D*, *MEF2C*, *MS4A* cluster (*MS4A3-MS4A6E*), *NME8*, *PICALM*, *PTK2B*, *SLC24A4*, *SORL1*, and *ZCWPW1* have been associated with Alzheimer's disease (AD) in large meta-analyses. We aimed to determine whether established AD-associated genes are associated with longitudinal cognitive decline by examining aggregate variation across these gene regions. In two single-sex cohorts of older, community-dwelling adults, we examined the association between SNPs in previously implicated gene regions and cognitive decline (age-adjusted person-specific cognitive slopes) using a Sequence Kernel Association Test (SKAT). In regions which showed aggregate significance, we examined the univariate association between individual

Electronic supplementary material The online version of this article (doi:10.1007/s11357-016-9885-2) contains supplementary material, which is available to authorized users.

J. Nettiksimmons (✉)
Department of Psychiatry, University of San Francisco - California, 4150 Clement Street, Box VAMC-116H, San Francisco, CA 94121, USA
e-mail: jasminen@gmail.com

G. Tranah
California Pacific Medical Center Research Institute, Department of Epidemiology and Biostatistics, University of California - San Francisco, Mission Hall: Global Health & Clinical Sciences Building, 550 16th Street, 2nd floor, Box #0560, San Francisco, CA 94158-2549, USA
e-mail: GTranah@psg.ucsf.edu

D. S. Evans
California Pacific Medical Center Research Institute, Mission Hall: Global Health & Clinical Sciences Building, 550 16th Street, 2nd floor, Box #0560, San Francisco, CA 94158-2549, USA
e-mail: DEvans@psg.ucsf.edu

J. S. Yokoyama
Memory and Aging Center, University of California - San Francisco, Sandler Neurosciences Center, 675 Nelson Rising Lane, Suite 190, San Francisco, CA, USA
e-mail: jyokoyama@memory.ucsf.edu

K. Yaffe
Departments of Psychiatry, Neurology, and Epidemiology and Biostatistics, University of California - San Francisco, San Francisco Veterans Affairs Medical Center, 4150 Clement Street, Box 181, San Francisco, CA 94121, USA
e-mail: kristine.yaffe@ucsf.edu

SNPs in the region and cognitive decline. Only two of the original AD-associated SNPs were significantly associated with cognitive decline in our cohorts. We identified significant aggregate-level associations between cognitive decline and the gene regions *BIN1*, *CD33*, *CELF1*, *CR1*, *HLA* cluster, and *MEF2C* in the all-female cohort and significant associations with *ABCA7*, *HLA* cluster, *MS4A6E*, *PICALM*, *PTK2B*, *SLC24A4*, and *SORL1* in the all-male cohort. We also identified a block of eight correlated SNPs in *CD33* and several blocks of correlated SNPs in *CELF1* that were significantly associated with cognitive decline in univariate analysis in the all-female cohort.

Keywords SNP associations · Candidate AD genes · Cognitive decline

Introduction

Single nucleotide polymorphisms (SNPs) have been identified as risk/protective factors for Alzheimer's disease (AD). Variants in the well-studied *APOE* gene have the strongest known genetic association with AD. In addition to *APOE*, SNPs in and near *ABCA7*, *BIN1*, *CASS4*, *CD2AP*, *CD33*, *CELF1*, *CLU*, *complement receptor 1* (*CR1*), *EPHA1*, *EXOC3L2*, *FERMT2*, *HLA cluster (DRB5-DQA)*, *INPP5D*, *MEF2C*, *MS4A* cluster (*MS4A3-MS4A6E*), *NME8*, *PICALM*, *PTK2B*, *SLC24A4*, *SORL1*, and *ZCWPW1* have been identified in large meta-analyses as being associated with AD (Harold et al. 2009; Hollingworth et al. 2011; Lambert et al. 2013; Naj et al. 2011; Seshadri et al. 2010). In general, the effect sizes for individual SNPs, other than those in *APOE*, are modest (odds ratios <1.25) and the extent to which these variants are causally linked to AD is still unclear.

While genome-wide association studies (GWAS) have been valuable platforms for identifying candidates for disease-related genetic variants, there are several limitations to these large-scale association studies. Most notably, the vast number of statistical comparisons being performed without a priori hypotheses requires corrections for multiple comparisons, thereby setting a stringent requirement to achieve statistical significance. In addition, large-scale GWAS have generally not examined additive or multiplicative SNP interactions within genomic regions. This is particularly problematic as collections of variants within genomic regions do not work in isolation and likely have a combined influence on phenotypic expression, particularly for complex phenotypes such as AD and cognitive decline (Schork et al. 2009; Torkamani et al. 2008).

To address some of the limitations encountered with univariate SNP-based analysis, novel methods have been developed to examine groups of SNPs in aggregate. Aggregate SNP methods reduce the total number of tests performed and increase power by taking advantage of the linkage disequilibrium (LD) across multiple SNPs (Wu et al. 2010). By reducing the number of tests and increasing power, smaller sample sizes are required compared with traditional GWAS. The Sequencing Kernel Association Test (SKAT) is one such method that can simultaneously test the association between multiple SNPs and a single outcome while controlling for covariates (Wu et al. 2011). Significance from SKAT analysis for a particular gene provides additional evidence for the involvement of the gene in the phenotype and provides a more focused target to search for functional variation contributing to underlying biology.

We set out to determine whether established AD-associated genes are associated with longitudinal cognitive decline by examining aggregate variation across these gene regions in two populations of older adults. This analysis posits that variation across an entire gene region—rather than a single SNP in isolation—plays a role in cognitive decline.

Methods

Participants

We examined two longitudinal cohorts of Caucasian older adults, one male (Osteoporotic Fractures in Men—MrOS) and one female (Study of Osteoporotic Fractures—SOF). During the MrOS baseline examination from 2000 to 2002, 5994 community-dwelling men 65 years or older were enrolled at six clinical centers in the USA: Birmingham, Alabama; Minneapolis, Minnesota; Palo Alto, California; Pittsburgh, Pennsylvania; Portland, Oregon; and San Diego, California as previously described (Blank et al. 2005; Orwoll et al. 2005). During the SOF baseline examination from 1986 to 1988, 9704 community-dwelling white women 65 years or older were enrolled from population-based listings in four areas of the USA: Baltimore, Maryland; Minneapolis, Minnesota; Pittsburgh, Pennsylvania; and Portland,

Oregon as previously described (Cummings et al. 1990). For both studies, individuals were not eligible to participate if they reported bilateral hip replacement or required the assistance of another person in ambulation at the baseline examination. Written informed consent was obtained from all SOF and MrOS participants, and the Institutional Review Board at each study site approved the study. Only participants with at least two cognitive testing time points were retained (SOF $N = 3267$, MrOS $N = 3026$).

Measures

For each cohort, we examined cognitive decline over 10 years of follow-up using a global cognitive test (abbreviated Mini-Mental Status Examination (MMSE) in SOF, Modified Mini-Mental Examination (3MS) in MrOS); each test was administered up to four times. Cognitive decline (person-specific slope) was calculated using linear mixed-effects regression models adjusted for age (Gould et al. 2001). Genotyping, imputation, and quality controls were performed identically in both cohorts using the Illumina HumanOmni1_Quad_v1-0 B array. Genotypes were called using the Illumina's BeadStudio calling algorithm. The sample quality control exclusion criteria were sample call rate <97 %, excessive autosomal heterozygozity, first and second degree relatives, genotypic sex mismatch using X and Y chromosome probe intensities, and gross chromosome abnormalities. Genotyped SNPs with GenTrain scores <0.6, cluster separation scores <0.4, call rates <97 %, or minor allele frequency (MAF) <0.01 were excluded. Also, autosomal SNPs with Hardy-Weinberg Equilibrium (HWE) p value <10–4 were excluded and genotype clusters for SNPs on chrX, chrY, chrXY, and chrMT (mitochondrial chromosomes) were reviewed manually. Autosomal SNPs (714,543) passed quality control. Imputation was done using MaCH (v 1.0.17, phasing) (Li and Abecasis 2006) and Minimac v 2011-08-12 beta (Howie et al. 2012) for HapMap phase II release 22 build 36, oriented on the positive strand. A combined panel of CEU, YRI, CHB, and JPT HapMap samples was used as a reference panel for the consensus-phased haplotypes.

Statistical analyses

First, we evaluated the relationship between the previously identified "sentinel" SNPs associated with AD and cognitive decline (person-specific age-adjusted cognitive slopes as outcome) in our cohorts using linear regression adjusting for the first four cohort-specific principal components. The sentinel SNPs we considered achieved genome-wide significance in recent, large meta-analyses. Then, we evaluated the associations between cognitive decline and candidate genes across multiple SNPs simultaneously using the SKAT package in R with a MAF cutoff of ≥2 % for SNP inclusion in this analysis (Wu et al. 2011). We considered multiple kernel functions: weighted and unweighted versions of the linear kernel and the identity-by-state (IBS) kernel. The linear kernel models each SNP as a variable in a regression model having its own beta coefficient. The IBS kernel models the aggregate variation using genetic similarity between individuals. The weighted kernel functions upweight rarer variants, whereas all SNPs are treated the same in the unweighted version (Wu et al. 2010). We also considered a kernel function which allows for multiplicative SNP interactions within the gene region. The different kernel functions for a given gene region and sample are not independent tests but may provide clues to the structure or relationship of the SNPs underpinning the association. For each gene, we included all available SNPs 30 kb upstream and downstream from the gene boundaries (RefSeq NCBI36/hg18), with two exceptions: *NME8* for which we went 47 kb downstream and *PICALM* in which we went 88 kb upstream in order to capture the sentinel SNPs. The first four principal components (cohort specific) were included in SKAT analyses to control for population stratification. For genes which showed a significant association, we performed a further SKAT analysis dropping the sentinel SNPs and all other SNPs in high LD with the sentinel SNPs ($R^2 \geq 0.80$ HapMap CEU from 1000 genomes). Gene-level p values for the primary SKAT analyses have not been adjusted for multiple comparisons as these analyses were based on a priori hypotheses. For gene regions exhibiting statistical significance (p value < 0.05) in the SKAT analysis after dropping the top associated SNPs, univariate analysis was performed on the remaining SNPs within the gene region in order to determine how additional SNPs at these loci contribute to cognitive decline (adjusted for first four cohort-specific principal components). P values for univariate analyses have been adjusted as noted in the results. Expression quantitative trait loci (eQTL) were examined for significant SNPs across ten human brain regions using the UK Brain Expression Consortium database (Ramasamy et al. 2014). Finally,

we performed permutation tests to identify significant enrichment of nominally significant results. In each cohort, the SKAT analysis was repeated 350 times with a different permutation of cognitive slopes as the outcome.

Because the two cohorts used different cognitive tests and differed in composition in several ways, including sex and study initiation date, we did not combine the cohorts. Rather, we performed stratified analyses and compared the results from both cohorts qualitatively. Although SOF had a longer follow-up at the time of analysis, having been initiated earlier, for the purpose of consistency, we have restricted follow-up for both studies to the first 10 years after baseline.

Results

The average baseline age of male participants (MrOS) was 73.4 years (standard deviation (SD) 5.7); average baseline age of female participants (SOF) was 71.0 years (SD 4.9) ($p < 0.001$). MrOS participants were more likely to have high education (56 % with a college degree vs 18 % in SOF, $p < 0.001$). SOF participants had a higher number of cognitive assessments with 74 % of SOF participants having had four cognitive assessments as compared to 21 % of MrOS participants with four assessments ($p < 0.001$). The average baseline cognitive assessment score in both cohorts was high (24.8 ± 1.5 out of a possible score of 26 in SOF and 94.4 ± 4.5 out of a possible score of 100 in MrOS).

We first tested each sentinel SNP available in our data for an association with cognitive decline (Table 1). Only two of the individual sentinel SNPs showed significant associations with cognitive decline in the female cohort (rs3764650 in *ABCA7*, $p = 0.01$; rs3865444 in *CD33*, $p = 0.01$). None of the top AD-associated SNPs were associated with cognitive decline in the male cohort. Ten of the 35 sentinel SNPs were not available in our data and could not be tested individually.

Next, we tested each gene region for aggregate association. Several gene regions showed significant aggregate associations with cognitive decline using one or more kernel functions: *BIN1*, *CD33*, *CELF1*, *CR1*, *HLA* cluster, and *MEF2C* in the female cohort and *ABCA7*, *HLA* cluster, *PICALM*, *PTK2B*, *SLC24A4*, and *SORL1* in the male cohort (Table 2). The significance was largely consistent regardless of whether the analytical method for assessing SNP structure/variation (kernel function) was linear or IBS. In contrast, most gene regions which showed any significance, showed significance using either weighted or unweighted methods, not both, which may provide some clue as to the rarity of the group of SNPs driving the significance. Two gene regions showed significance with the interaction kernel function *CELF1* in women and *SORL1* in men. We then tested each significant gene region for continued aggregate association with cognitive decline after removal of the sentinel SNP(s) which led us to these particular genomic regions of interest. After dropping the sentinel SNP(s) and all SNPs in LD with them ($R^2 \geq 0.80$ HapMap CEU), the aggregate significance in gene regions which were significant in the primary analysis remained largely unchanged. Resulting p values along with the number of SNPs dropped is shown in Table 2. Our data did not contain the sentinel SNP or any correlated SNPs for the HLA cluster, so the secondary analysis is identical to the primary analysis.

Next, we performed univariate analysis within significant gene regions to identify new candidate variants for cognitive decline. In all 11 gene regions which showed aggregate significance with cognitive decline, we found individual SNPs that were nominally significantly associated with cognitive decline in univariate analysis in their respective cohorts. After correction for multiple comparisons by gene region, only univariate associations in CD33 and *CELF1* in the female cohort remained significant.

Univariate analysis in *CD33* in the all-female cohort revealed a block of eight SNPs (rs273638, rs273639, rs1697553, rs2455069, rs12609179, rs1566576, rs1697573, rs273634) in strong LD which were individually significantly associated with greater cognitive decline (all unadjusted $p = 0.0001$, all $p = 0.001$ after adjustment for false discovery via Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) (shown in Supplementary Fig. 1). These SNPs were not associated with CD33 gene expression across ten brain regions (Sherry et al. 2001). Several of the *CD33* SNPS exhibit highly significant effects on hyaluronan synthase-1 (*HAS1*) expression in the temporal cortex, with the strongest effect for rs2455069 ($p = 1.4e\text{-}05$), which encodes an Arg69Gly substitution in CD33.

Also in the female cohort, we identified eight SNPs in *CELF1* which remained significant ($p = 0.002$–0.02) after adjustment for false discovery within the gene. Two blocks of SNPs (rs11604680, rs1317149, rs4752845 and rs7124681, rs7928842) were associated

Table 1 Gene regions examined

Gene	Chrom.	Bounds	No. MrOS SNPs	No. SOF SNPs	SNP ID	SOF *p* value	MrOS *p* value
ABCA7	19	961–1047	49	49	rs3764650	**0.01**	0.60
					rs4147929	–	–
BIN1	2	127,492–127,611	125	126	rs744373	0.84	0.17
					rs6733839	–	–
					rs7561528	0.46	0.24
CASS4	20	54,391–54,498	105	105	rs7274581	–	–
CD2AP	6	47,523–47,733	131	131	rs9349407	0.43	0.48
					rs10948363	0.33	0.77
CD33	19	56,390–56,465	80	69	rs3865444	**0.01**	0.35
CELF1	11	47,414–47,561	44	44	rs10838725	–	–
CLU	8	27,480–27,558	75	74	rs11136000	0.18	0.68
					rs9331896	0.16	0.56
					rs1532278	0.16	0.66
CR1	1	205,706–205,912	133	133	rs3818361	0.99	0.13
					rs6656401	0.84	0.08
					rs6701713	0.95	0.12
EPHA1	7	142,768–142,846	34	33	rs11767557	0.99	0.52
					rs11771145	0.56	0.45
EXOC3L2	19	50,378–50,459	38	38	rs597668	0.87	0.96
FERMT2	14	52,364–52,518	120	120	rs17125944	0.58	0.94
HLA cluster (DRA-DQB1)	6	32,486–32,772	305	304	rs9271192	–	–
INPP5D	2	233,603–233,811	160	160	rs35349669	–	–
MEF2C	5	88,020–88,266	135	137	rs190982	0.06	0.35
MS4A cluster (MS4A3-MS4A6E)	11	59,551–59,895	345	343	rs670139	0.87	0.45
					rs610932	0.76	0.41
					rs4938933	–	–
					rs983392	–	–
NME8	7	37,808–37,937	227	226	rs2718058	–	–
PICALM	11	85,316–85,547	210	210	rs3851179	0.75	0.64
					rs10792832	0.74	0.68
					rs561655	0.90	0.40
PTK2B	8	27,195–27,403	238	236	rs28834970	–	–
SLC24A4	14	91,829–92,068	229	228	rs10498633	0.80	0.66
SORL1	11	120,798–121,040	175	172	rs11218343	0.82	0.38
ZCWPW1	7	99,806–99,894	26	26	rs1476679	0.11	0.07

SNPs 30 kb up/downstream from NCBI36/hg18 gene boundaries were included with the exception of NME8 for which we included 47 kb downstream and PICALM in which we included 88 kb upstream in order to capture the sentinel SNP(s). *P* values reflect the strength of association between named SNP and cognitive slope in each cohort. SNPs without *p* values were not available in our data

Chrom. chromosome, *Bounds* start and end position used in SKAT analysis

Significant ($p < 0.05$) *p*-values are presented in bold

with better cognitive performance; another block of three SNPs (rs2242081, rs10742814, rs11039280) were associated with worse cognitive performance. The rs11604680 and rs1317149 SNPs from the first block are located 3′ of *CELF1*. Each of these SNPs was associated with *CELF1* expression in the hippocampus

Table 2 *P* values for aggregate SNP association with cognitive decline by gene region in SOF and MrOS

		All available SNPs in gene region					Excludes sentinel SNPs (and LD $R^2 \geq 0.8$)					
Gene	Cohort	Linear	IBS	W. Linear	W. IBS	Interaction	Linear	IBS	W. Linear	W. IBS	Interaction	No. SNPs dropped
ABCA7	SOF	0.18	0.12	0.64	0.59	0.19						
	MROS	0.17	**0.04**	**0.04**	**0.047**	0.12	0.16	**0.04**	**0.04**	**0.05**	0.11	2
BIN1	SOF	0.57	0.68	**0.03**	**0.03**	0.36	0.55	0.66	**0.03**	**0.03**	0.36	3
	MROS	0.34	0.36	0.44	0.40	0.23						
CASS4	SOF	0.10	0.17	0.23	0.57	0.07						
	MROS	0.27	0.39	0.36	0.80	0.31						
CD2AP	SOF	0.69	0.58	0.22	0.19	0.81						
	MROS	0.80	0.92	0.71	0.42	0.79						
CD33	SOF	**0.01**	**0.02**	0.24	0.23	0.06	**0.02**	**0.02**	0.24	0.23	0.07	2
	MROS	0.65	0.86	0.46	0.83	0.56						
CELF1	SOF	**0.004**	**0.01**	0.50	0.56	**0.01**	**0.003**	**0.01**	0.50	0.56	**0.01**	3
	MROS	0.41	0.59	0.14	0.40	0.45						
CLU	SOF	0.85	0.81	0.93	0.92	0.91						
	MROS	0.26	0.25	0.15	0.21	0.21						
CR1	SOF	0.51	0.46	**0.005**	**0.003**	0.52	0.48	0.43	**0.005**	**0.003**	0.49	18
	MROS	0.26	0.21	0.35	0.35	0.36						
EPHA1	SOF	0.55	0.24	0.18	0.17	0.52						
	MROS	0.47	0.87	0.69	0.75	0.45						
EXOC3L2	SOF	0.17	0.17	0.60	0.63	0.14						
	MROS	0.50	0.48	0.18	0.22	0.73						
FERMT2	SOF	0.18	0.24	0.38	0.40	0.17						
	MROS	0.80	0.86	0.61	0.80	0.73						
HLA	SOF	0.13	0.18	**0.03**	**0.02**	0.11	0.13	0.18	**0.03**	**0.02**	0.11	0
	MROS	0.30	0.28	**0.02**	**0.03**	0.16	0.30	0.28	**0.02**	**0.03**	0.16	0
INPP5D	SOF	0.63	0.78	0.86	0.95	0.68						
	MROS	0.19	0.32	0.76	0.52	0.50						
MEF2C	SOF	0.06	**0.03**	0.36	0.51	0.12	0.06	**0.03**	0.36	0.51	0.12	2
	MROS	0.74	0.66	0.94	0.68	0.28						
MS4A	SOF	0.99	0.41	0.95	0.95	0.56						
	MROS	0.75	0.84	0.23	0.23	0.45						
NME8	SOF	0.76	0.82	0.50	0.47	0.80						
	MROS	0.90	0.95	0.87	0.88	0.91						
PICALM	SOF	0.73	0.74	0.68	0.79	0.42						
	MROS	0.08	**0.04**	**0.02**	**0.04**	0.12	0.07	**0.04**	**0.02**	**0.04**	0.08	14
PTK2B	SOF	0.62	0.73	0.79	0.43	0.80						
	MROS	0.79	0.81	**0.03**	**0.04**	0.81	0.79	0.81	**0.03**	**0.04**	0.82	1
SLC24A4	SOF	0.90	0.97	0.70	0.69	0.84						
	MROS	0.40	0.56	0.06	**0.03**	0.59	0.39	0.55	0.06	**0.03**	0.59	2
SORL1	SOF	0.32	0.40	0.10	0.07	0.32						
	MROS	**0.04**	**0.04**	0.69	0.73	**0.03**	**0.04**	**0.04**	0.70	0.74	**0.03**	1
ZCWPW1	SOF	0.23	0.13	0.56	0.46	0.52						
	MROS	0.33	0.24	0.70	0.63	0.60						

W weighted

[a] For the calculation of the "Exclude sentinel SNPs" *p* value, sentinel SNP along with any SNPs which were in high linkage disequilibrium with sentinel SNP ($R^2 \geq 0.8$) were dropped

Significant ($p < 0.05$) *p*-values are presented in bold

(rs1317149, $p = 0.025$, and rs4752845, $p = 0.015$). The remaining SNPs are intronic within *CELF1* and are associated with gene expression CELF1 in the thalamus. The rs7124681 and rs7928842 SNPs from the second SNP block that were associated with improved cognitive performance exhibited lower *CELF1* expression with each copy of the minor allele (rs7124681, $p = 0.022$, and rs7928842, $p = 0.013$). The three SNPs from block three that were associated with poor cognitive function exhibited higher *CELF1* expression with each copy of the minor allele (rs2242081 and rs10742814, $p = 0.02$, and rs11039280, $p = 0.027$). Several of the *CELF1* SNPS exhibit highly significant effects on *MTCH2* expression in the cerebral cortex, with the strongest effect for rs7928842 ($p = 8.1e-07$).

Finally, we performed permutation tests to identify significant enrichment of nominally significant associations across multiple gene regions. We focused the permutation analyses on the linear and weighted linear kernel functions since the linear and IBS kernel functions performed similarly in the original analyses. We considered a gene region "significant" if the p value for either of the kernel functions was <0.05. In our main analysis, we found significant results for six gene regions in each cohort. The probability of finding significant results for six or more gene regions in the female cohort was 0.006, and the probability of six or more significant gene regions in the male cohort was 0.01.

Discussion

We examined gene regions based on previously identified AD-associated SNPs using an aggregate testing method (SKAT) and found that several regions were associated with cognitive decline in each cohort. Among SOF women, *BIN1*, *CD33*, *CELF1*, *CR1*, *HLA* cluster, and *MEF2C1* were significantly associated with cognitive decline and remained so even after removing the sentinel SNPs and all other SNPs in high LD ($R^2 \geq 0.8$, HapMap CEU) with the sentinels. Among MrOS men, *ABCA7*, *HLA* cluster, *PICALM*, *PTK2B*, *SLC24A4*, and *SORL1* were significantly associated with cognitive decline and remained significant after dropping the sentinel SNP and SNPs in LD with the sentinel SNP.

It is perhaps not surprising that dropping the top AD-associated SNPs did not alter the gene-level aggregate since the individual AD-associated SNPs were largely not associated with cognitive decline in these cohorts. In fact, *CD33* in SOF was the only gene region that had both a significant AD-associated SNP and showed a significant aggregate association with cognitive decline.

The fact that some gene regions exhibited aggregate associations using only kernel functions which upweight rarer SNPs (*BIN1*, *CR1*, and *HLA* cluster in females; *HLA* cluster and *PTK2B* in males) and others only showed association using unweighted kernel functions (*CD33* and *CELF1* in females; *SORL1* in males) likely reflects the SNP characteristics underlying the gene-level associations. The SNPs which were nominally significant in univariate analysis in *BIN1* and *CR1* in the female cohort had relatively rare minor alleles (MAF 0.02–0.07). In contrast, in *CD33* and *CELF1*, which were only significant in the female cohort using unweighted kernel functions, the nominally significant SNPs had a much higher MAF (0.06–0.44, with most SNPs in the 0.31–0.44 range). A similar result occurred in the all-male cohort: in nominally significant SNPs in *PTK2B*, which showed weighted aggregate significance, the minor alleles were less common (MAF 0.02–0.16); the MAF was higher in nominally significant SNPs from *SORL1*, which showed unweighted aggregate significance (0.04–0.50, with MAF for most SNPs >0.20). The *HLA* cluster was a curious departure from this pattern in both cohorts. Despite having aggregate significance in both cohorts using only kernel functions which upweight rarer SNPs, only 11 % of the nominally significant SNPs in the female cohort and 40 % of the nominally significant SNPs in the male cohort had MAF ≤ 0.05. The reason for this result is unclear, although none of the SNPs were strongly associated enough to remain significant after adjustment.

In addition to a significant aggregate association with *CD33*, we found eight SNPs forming an LD block in *CD33* that were also significantly associated with greater cognitive decline in univariate analysis. CD33 is a transmembrane glycoprotein, a member of the sialic acid-binding immunoglobulin-like lectins (Siglecs). *CD33* is known to perform a number of functions including cell–cell communication inhibiting immune response (Crocker et al. 2012; Pillai et al. 2012), immune cell growth, adhesion processes in immune or malignant cells, and endocytosis (Crocker et al. 2007; von Gunten and Bochner 2008). Recent research has focused on understanding how *CD33* may modify AD susceptibility or disease course. *CD33* expression is modestly increased in AD, and the previously identified protective minor allele of the top AD-associated SNP (rs3865444)

has been associated with reductions in *CD33* microglia expression (Griciuc et al. 2013; Malik et al. 2013). Microglia are hypothesized to play a role in phagocytosis of amyloid-beta, a primary pathological protein in AD, and recent research has borne out an association between the protective allele and increased uptake of amyloid-beta, as well as reduced amyloid plaque burden and insoluble amyloid-beta levels (Bradshaw et al. 2013; Griciuc et al. 2013). In our study, the large eight-SNP LD block in *CD33* exhibiting significant univariate associations with cognitive decline is located within the 5′ end of the gene region including the promoter, suggesting it may affect expression levels or protein function. However, these SNPs did not impact *CD33* expression across several brain regions in a large eQTL database. It is possible that these SNPs impact expression of nearby genes and not *CD33*. Indeed, several of these SNPs, including rs2455069 encoding the Arg69Gly substitution in *CD33*, were associated with strong effects on *HAS1* gene expression in the temporal cortex, a region of the brain affected in AD. While the *HAS1* gene is approximately 500 kb away from rs2455069, Ramasamy et al. (2014) observed that numerous brain eQTL signals were within 1 Mb of their target gene and often acted heterogeneously among genomic regions and exons. Hyaluronan (encoded by HAS1) transcription increases with aging. *HAS1* is transcriptionally upregulated in astrocytes during normative aging and is linked to the accumulation of the hyaluronan in gray matter where it potentially inhibits astrogliosis and limits oligodendrocyte progenitor cell maturation (Cargill et al. 2012). Further study is required to identify the function of rs2455069 and whether its effects on neurodegeneration are mediated through *CD33* or another gene (e.g., *HAS1*).

In addition to the strong aggregate result in *CELF1* in the female cohort, we also found eight SNPs which were significantly associated with greater cognitive decline in the *CELF1* region. The *CELF* family of proteins are involved in the regulation of RNA processing including pre-mRNA alternative splicing, RNA editing, deadenylation, mRNA stability, and translation. *CELF* proteins have been implicated in a number of disease including AD, potentially through the regulation of tau protein aggregates (Dasgupta and Ladd 2012). Interestingly, the two *CELF1* SNPs that were associated with improved cognitive performance exhibited lower *CELF1* expression levels in the thalamus and the three *CELF1* SNPs that were associated with poor cognitive function exhibited higher *CELF1* expression in the thalamus. While these results suggest that the *CELF1* SNPs impacting cognition may be modifying expression levels in the thalamus, it is possible that these SNPs are also impacting expression of nearby genes. Several of these SNPs are highly associated with *MTCH2* expression levels in the cerebral cortex. *MTCH2* is located approximately 100 kb away from rs7928842, the SNP with the strongest effect on *MTCH2* expression. *MTCH2* encodes mitochondrial carrier 2 which likely plays a role in cellular apoptosis (PMID nos. 18614015 and 15899861) and has been associated with obesity (PMID no. 21795451).

Although the sentinel SNP from *CR1* was not significantly associated with cognitive decline in either sample, *CR1* showed the strongest aggregate association with the weighted kernel functions in SOF. We found nine nominally significant SNPs, but none were significant after adjusting for multiple comparisons. CR1 is one of 30 proteins that make up the complement system, which participate in the regulation of inflammation and immune reaction. CR1 is expressed in the brain and may play a role in amyloid-beta clearance (Crehan et al. 2012).

To our knowledge, this analysis is the first to look at aggregate-level genetic associations with cognitive decline. Using previously identified AD-associated SNPs as sentinels, we targeted the entire gene region to assess whether there was aggregate association with cognitive decline, with and without the sentinel SNP. This approach allowed us to test whether variation across the entire gene region impacts cognitive decline. Both SOF and MrOS are community-based samples rather than clinical samples, giving our results greater generalizability. Another strength of our analysis was the ability to follow participants longitudinally to capture cognitive decline rather than simply observing cognitive status at a single time point, which can be subject to greater confounding. Although several of the sentinel SNPs and gene regions were associated with cognitive decline in each cohort, it is unclear why the results in the two cohorts differed. The most obvious difference between the cohorts is in their single-sex composition. There may be biological differences in risk of cognitive decline due to sex-related differences in hormones, immune regulation, inflammatory response, and comorbidities. There were also differences in cognitive testing: SOF used a less sensitive global cognitive test (MMSE); however, the participants were more likely to have had four

testing time points. Thus, cognitive change observed in SOF may have been quite pronounced. By contrast, the MrOS study used a more sensitive global cognitive test (3MS), but many of the participants only had two testing time points, which may not have been adequate to observe substantial change or may have created noise by identifying inconsequential change. This analysis had several limitations. First, we were unable to test several of the sentinel SNPs because they were unavailable in our data. Second, the candidate genes we selected were motivated by AD-associated SNPs, whereas we have examined cognitive decline, which likely results from multiple pathologies, not just AD. AD is the most common form of dementia, and a large proportion of the cognitive decline that we observed in these cohorts is probabilistically due to AD (Barker et al. 2002). But the fact that the top AD-associated SNPs were largely unassociated with cognitive decline in both cohorts might suggest that AD-targeted regions do not translate well to general cognitive decline. Finally, the MMSE is not the most sensitive test for capturing early cognitive decline but more sensitive cognitive tests were unavailable in the early years of the SOF study.

Many of the sentinel SNPs and associated gene regions were not associated with cognitive decline in either sample. This finding is supported by two recent studies in other cohorts which found limited or no association between risk genes for AD and cognitive aging (Harris et al. 2014; Verhaaren et al. 2013). One potential explanation for the lack of consistent associations is that the gene regions suggested by the top AD SNPs were not causally associated with AD, and the associated SNPs are in LD with the causal SNPs in a different genomic region. Another possibility is that the AD-associated sentinel SNPs which informed our analysis are AD specific and other pathologies such as vascular disease have stronger contributions to the cognitive decline in these cohorts. A recent study found evidence of synergistic interaction effects among AD-associated genes and the development of AD (Ebbert et al. 2014). Thus, an association between AD risk genes and cognitive decline may involve more complicated relationships than have been tested here. Another possible explanation is that our study was insufficiently powered.

In conclusion, we identified aggregate-level associations between cognitive decline and the gene regions *ABCA7*, *BIN1*, *CD33*, *CELF1*, *CR1*, *HLA*, *MEF2C*, *PICALM*, *PTK2B*, *SLC24A4*, and *SORL1*. In addition to the gene-level association, novel *CD33* and *CELF1* SNP associations independent of previously identified sentinel SNPs were identified for cognitive decline. If replicated in independent population-based studies of cognitive function, our results suggest that *CD33* and *CELF1* may be important targets for functional follow-up studies related to cognitive aging.

Acknowledgments The Study of Osteoporotic Fractures (SOF) is supported by the National Institutes of Health funding. The National Institute on Aging (NIA) provides support under the following grant numbers: R01 AG005407, R01 AR35582, R01 AR35583, R01 AR35584, R01 AG005394, R01 AG027574, and R01 AG027576. The Osteoporotic Fractures in Men (MrOS) Study is supported by the National Institutes of Health funding. The following institutes provide support: the National Institute on Aging (NIA), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the National Center for Advancing Translational Sciences (NCATS), and NIH Roadmap for Medical Research under the following grant numbers: U01 AG027810, U01 AG042124, U01 AG042139, U01 AG042140, U01 AG042143, U01 AG042145, U01 AG042168, U01 AR066160, and UL1 TR000128. The NIAMS provides funding for the MrOS ancillary study "GWAS in MrOS and SOF" under the grant number RC2 AR058973. TheNIAMS provides funding for the MrOS ancillary study "Replication of candidate gene associations and bone strength phenotype in MrOS" under the grant number R01 AR051124. Dr. Yaffe is supported in part by a National Institute of Aging Grant (K24AG031155). Dr. Yokoyama is supported in part by Larry L. Hillblom Foundation 2012-A-015-FEL, National Institute on Aging K01 AG049152 and Diversity Supplement to P50 AG023501, and AFTD Susan Marcus Memorial Fund Clinical Research Grant.

References

Barker WW, Luis CA, Kashuba A, Luis M, Harwood DG, Loewenstein D, … Duara R (2002) Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. Alzheimer Dis Assoc Disord 16(4): 203–212.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57:289–300

Blank JB, Cawthon PM, Carrion-Petersen ML, Harper L, Johnson JP, Mitson E, Delay RR (2005) Overview of recruitment for the osteoporotic fractures in men study (MrOS). Contemp Clin Trials 26(5):557–568. doi:10.1016/j.cct.2005.05.005

Bradshaw EM, Chibnik LB, Keenan BT, Ottoboni L, Raj T, Tang A, … De Jager PL (2013) CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. Nat Neurosci 16(7):848–850. doi:10.1038/nn.3435

Cargill R, Kohama SG, Struve J, Su W, Banine F, Witkowski E, … Sherman LS (2012) Astrocytes in aged nonhuman primate brain gray matter synthesize excess hyaluronan. Neurobiol

Aging 33(4):830. e813–824. doi:10.1016/j.neurobiolaging.2011.07.006

Crehan H, Holton P, Wray S, Pocock J, Guerreiro R, Hardy J (2012) Complement receptor 1 (CR1) and Alzheimer's disease. Immunobiology 217(2):244–250. doi:10.1016/j.imbio.2011.07.017

Crocker PR, McMillan SJ, Richards HE (2012) CD33-related siglecs as potential modulators of inflammatory responses. Ann N Y Acad Sci 1253:102–111. doi:10.1111/j.1749-6632.2011.06449.x

Crocker PR, Paulson JC, Varki A (2007) Siglecs and their roles in the immune system. Nat Rev Immunol 7(4):255–266. doi:10.1038/nri2056

Cummings SR, Black DM, Nevitt MC, Browner WS, Cauley JA, Genant HK et al (1990) Appendicular bone density and age predict hip fracture in women. The Study of Osteoporotic Fractures Research Group. JAMA 263(5):665–668

Dasgupta T, Ladd AN (2012) The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. Wiley Interdiscip Rev RNA 3(1):104–121. doi:10.1002/wrna.107

Ebbert MT, Ridge PG, Wilson AR, Sharp AR, Bailey M, Norton MC, …. Kauwe JS (2014) Population-based analysis of Alzheimer's disease risk alleles implicates genetic interactions. Biol Psychiatry 75(9):732–737. doi: 10.1016/j.biopsych.2013.07.008

Gould R, Abramson I, Galasko D, Salmon D (2001) Rate of cognitive change in Alzheimer's disease: methodological approaches using random effects models. J Int Neuropsychol Soc 7(7):813–824

Griciuc A, Serrano-Pozo A, Parrado AR, Lesinski AN, Asselin CN, Mullin K, … Tanzi RE (2013) Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. Neuron 78(4):631–643. doi: 10.1016/j.neuron.2013.04.014

Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML,…. Williams J (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 41(10):1088–1093. doi: 10.1038/ng.440

Harris SE, Davies G, Luciano M, Payton A, Fox HC, Haggarty P, … Deary IJ (2014) Polygenic risk for Alzheimer's disease is not associated with cognitive ability or cognitive aging in non-demented older people. J Alzheimers Dis 39(3):565–574. doi: 10.3233/JAD-131058

Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, … Williams J (2011) Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 43(5):429–435. doi: 10.1038/ng.803

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44(8):955–959. doi:10.1038/ng.2354

Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, …. Amouyel P (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45(12):1452–1458. doi: 10.1038/ng.2802

Li Y, Abecasis GR (2006) Mach 1.0: rapid haplotype reconstruction and missing genotype inference. Am J Hum Genet S 79(3):2290

Malik M, Simpson JF, Parikh I, Wilfred BR, Fardo DW, Nelson PT, Estus S (2013) CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing. J Neurosci 33(33):13320–13325. doi:10.1523/JNEUROSCI.1224-13.2013

Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, … Schellenberg GD (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 43(5):436–441. doi: 10.1038/ng.801

Orwoll E, Blank JB, Barrett-Connor E, Cauley J, Cummings S, Ensrud K, … Stone K (2005) Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study—a large observational study of the determinants of fracture in older men. Contemp Clin Trials 26(5):569–585. doi: 10.1016/j.cct.2005.05.006

Pillai S, Netravali IA, Cariappa A, Mattoo H (2012) Siglecs and immune regulation. Annu Rev Immunol 30:357–392. doi:10.1146/annurev-immunol-020711-075018

Pruim R, Welch R, Sanna S, Teslovich T, Chines P, Gliedt T, … Willer C (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics (Oxford, England) 26(18):2336–2337.

Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, … Weale ME (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 17(10):1418–1428. doi: 10.1038/nn.3801

Schork NJ, Murray SS, Frazer K a, Topol EJ (2009) Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19:212–219. doi:10.1016/j.gde.2009.04.010

Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, …. Consortium E (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 303(18):1832–1840. doi: 10.1001/jama.2010.574

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29(1):308–311

Torkamani A, Topol EJ, Schork NJ (2008) Pathway analysis of seven common diseases assessed by genome-wide association. Genomics 92(5):265–272. doi:10.1016/j.ygeno.2008.07.011

Verhaaren BF, Vernooij MW, Koudstaal PJ, Uitterlinden AG, van Duijn CM, Hofman A, … Ikram MA (2013) Alzheimer's disease genes and cognition in the nondemented general population. Biol Psychiatry 73(5):429–434. doi: 10.1016/j.biopsych.2012.04.009

von Gunten S, Bochner BS (2008) Basic and clinical immunology of Siglecs. Ann N Y Acad Sci 1143:61–82. doi:10.1196/annals.1443.011

Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X (2010) Powerful SNP-set analysis for case–control genome-wide association studies. Am J Hum Genet 86(6):929–942. doi:10.1016/j.ajhg.2010.05.002

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89(1):82–93. doi:10.1016/j.ajhg.2011.05.029