Published in final edited form as:

Am J Med Genet B Neuropsychiatr Genet. 2009 June 5; 150B(4): 520–526. doi:10.1002/ajmg.b.30851.

Association Studies of 22 Candidate SNPs with Late-Onset Alzheimer's Disease

Jessica A. Figgins¹, Ryan L. Minster¹, F. Yesim Demirci¹, Steven T. DeKosky², and M. Ilyas Kamboh^{1,*}

¹Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA 15261

²Department of Neurology, University of Pittsburgh, Pittsburgh, PA 15213

Abstract

Alzheimer's disease (AD) is a complex and multifactorial disease with the possible involvement of several genes. With the exception of the *APOE* gene as a susceptibility marker, no other genes have been shown consistently to be associated with late-onset AD (LOAD). A recent genome-wide association study of 17,343 gene-based putative functional single nucleotide polymorphisms (SNPs) found 19 significant variants, including 3 linked to *APOE*, showing association with LOAD (Hum Mol Genet 2007; 16:865–873). We have set out to replicate the 16 new significant associations in a large case-control cohort of American Whites. Additionally, we examined six variants present in positional and/or biological candidate genes for AD. We genotyped the 22 SNPs in up to 1,009 Caucasian Americans with LOAD and up to 1,010 age-matched healthy Caucasian Americans, using 5' nuclease assays. We did not observe a statistically significant association between the SNPs and the risk of AD, either individually or stratified by *APOE*. Our data suggest that the association of the studied variants with LOAD risk, if it exists, is not statistically significant in our sample.

Keywords

Alzheimer's disease; genetics; age-at-onset; disease duration

Introduction

Alzheimer's disease (AD) is the most common form of dementia in the elderly, and has a significant impact on public health in the United States. Although AD has been described for more than 100 years, no effective treatment has yet been developed. Late-onset AD (LOAD) is a complex disorder in which gene-gene as well as gene-environment interactions are involved in the etiology of the disease. Although the exact etiology of LOAD has not been identified, first-degree relatives of individuals with LOAD have been found to have an increased risk of developing the disease. It is estimated that genetic factors explain up to 79% of the risk for LOAD (Gatz et al., 2006). Despite the evidence of substantial genetic effect on LOAD, to date the only known and significant genetic risk factor is *APOE*. The reason researchers are still looking for additional genetic factors for LOAD is because the *APOE* itself does not explain the entire heritability of LOAD and only certain percentage of LOAD cases carry an *APOE* risk allele.

*Corresponding Author: M. Ilyas Kamboh, Ph.D., Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261; Telephone: 412-624-3066; Fax: 412-383-7844; kamboh@pitt.edu.

In a genome-wide association study examining 17,343 putative functional SNPs located in 11,221 unique genes, 19 SNPs were found to have a statistically significant association with LOAD (Grupe et al., 2007). Three of these SNPs were located near the *APOE* gene and were in linkage disequilibrium (LD) with the known *APOE* polymorphism associated with LOAD. Our goal was to replicate the 16 new associations in a large case-control sample of white Americans (SNPs 1–16 in Table I). Additionally, we examined six positional and/or biological candidates SNPs as risk factors for AD (SNPs 17–22 in Table I). The secondary goal of the study was to examine the association of 22 SNPs with quantitative traits linked to LOAD, including age-at-onset (AAO), Mini-Mental State Examination (MMSE) score and disease duration.

Materials and Methods

Subjects were 2,019 Caucasian American individuals, including 1,009 LOAD cases (AAO $\geq$ 60 years) and 1,010 older controls (age $\geq$ 60 years). Cases were derived from the University of Pittsburgh Alzheimer's Disease Research Center (ADRC) that included 67.7% female with a mean AAO of $72.85 \pm$ (SD) 6.24 years. Clinical diagnosis of AD cases was made according to the NINCDS/ADRDA criteria. Our ADRC follows a standard evaluation protocol, including medical history, general medical and neurological examinations, a psychiatric interview, neuropsychological testing and MRI scan. Cognitively normal controls were recruited from the same geographical area as the cases. The mean age of controls at baseline was $74.07 \pm$ 6.20 years, and 59.8% were females. The subjects were recruited with informed consent, and the study was approved by the University of Pittsburgh Institutional Review Board.

DNA was isolated from blood using the QIAamp Blood DNA Maxi Kit protocol (Qiagen, Valencia CA), or from brain tissue using the QIAamp DNA Mini Kit tissue protocol (Qiagen). A small number of samples with a low amount of DNA were whole-genome amplified using the GenomiPhi kit (GE Healthcare, Piscataway NJ).

The genotypes were determined using TaqMan SNP Genotyping Assays and Genotyping Master Mix according to manufacture's protocol (Applied Biosystems, Foster City, CA). The TaqMan analysis was performed after all samples were placed on 384 well plates. Every plate had a mixture of cases and controls, and ten percent of the samples were repeated in order to assess error rate.

Allele and genotype frequencies were calculated by the direct allele-counting method. Goodness of fit to Hardy-Weinberg Equilibrium (HWE) was tested using the χ^2 test. Differences between genotype and allele frequencies in cases and controls were tested with the χ^2 or Fisher's exact tests as appropriate. These statistics were calculated using R 2.2.0 with the genetics package (R Development Core Team, 2005). Analysis of variance was performed to estimate the impact of genotyping on AAO, disease duration (age at death – age at diagnosis), and MMSE at baseline and change in MMSE score from the baseline to the latest score. Power to detect associations was determined with PS 2.1.30 (Dupont and Plummer, 1998).

Results

The genotyping error rate for all the SNPs but SNP 18 (rs2943634) was estimated to be $< 1\%$; the estimated error rate for SNP 18 was 2.8%. While the frequency of the *APOE**4 allele was significantly higher in cases than controls (33% vs. 11%; $p < 0.0001$) the frequency of the *APOE**2 allele was lower in cases than controls (3.5% vs. 8.7%; $p < 0.0001$).

Twenty-one SNPs were in HWE. However, SNP 2 (*PCK1*/rs8192708) was found to be out of HWE with a p value of 2.569×10^{-36} for cases and 1.401×10^{-29} for controls. A search through NCBI (dbSNP) found three groups reporting genotype and allele frequencies for this SNP.

These groups include HapMap (ss44233834), Perlegen (ss24451444) and a group from University of Washington (ss66859094). The allele frequencies in these studies, as well as in the study by Grupe et al. (2007) were similar to ours, but the genotype frequencies were not. The data from NCBI and the Grupe et al. (2007) were in HWE. When we examined the plots read from TaqMan, the clusters were distinct, and the calls were conservative. Each plate individually was also not within HWE, the magnitudes of the p values were similar for each of the plates. The cases and controls showed the same effect with similar p values. The pattern in all of the plates in cases and controls was an excess of both homozygotes and a shortage of heterozygotes in our data. We therefore investigated if there were any copy number variations in the region, none have been reported (UCSC Genome Browser).

The calculated genotype and allele frequencies for the 22 SNPs are listed in Table I. SNP 2 (*PCK1*/rs3745833), which was out of HWE, showed a statistical difference in allele frequency ($p = 0.015$) but not in genotype frequency ($p = 0.057$). On the other hand, SNP 7 (*UBD*/rs444013) showed a marginal difference in genotype distribution ($p = 0.048$) but not in allele frequency ($p = 0.168$). No other statistical significant association at $p < 0.05$ was observed. We used the risk alleles identified in the Grupe et al. (2007) paper and coded our population in order to determine how many risk alleles they carried in total (assuming a dosage model) and how many loci contained a risk allele (assuming a dominant model). We carried out a linear regression for each model adjusting for age, sex and *APOE* status. Analyses in either a dosage model ($p = 0.804$) or dominant model ($p = 0.634$) were not statistically significant.

No statistical interaction was observed between *APOE* and any of the SNPs examined. When analysis of variance was performed for MMSE baseline and change in score no significant associations were detected. However, analysis of variance for AAO showed significant associations with SNP 10, *MYH13*/rs2074877 ($p=0.00196$). *APOE* and gender-adjusted mean AAO for SNP 10 genotypes, CC, CT and TT were 72.65 ± 6.14 ($n = 389$), 72.50 ± 5.88 ($n = 448$) and 74.55 ± 6.27 ($n = 134$), respectively. Disease duration (age at death – age at onset) data were available on 80 AD cases. *APOE* and gender-adjusted disease duration showed significant association with SNP 13, *CTSS*/rs41271951 ($p = 0.006$) and SNP 14, *FAM63A*/rs41310885 ($p = 0.0014$). Genotype-specific mean disease duration for SNP 13 was 9.80 ± 3.81 years in the AA genotype ($n = 69$) and 6.44 ± 2.61 years for the AG genotype ($n = 11$). For SNP 14, adjusted mean disease duration was 9.78 ± 3.74 in the TT genotype ($n = 72$) and 5.37 ± 2.03 in the AT genotype ($n = 8$). Since both SNP 13 and SNP 14 are located in the same region on chromosome 1q21, we reasoned that the observed associations are due to linkage disequilibrium (LD). Indeed, there was a strong linkage disequilibrium between the two SNPs ($D' = 0.908$; $r^2 = 0.693$).

Discussion

We did not observe a statistically significant association between any of the 22 SNPs examined and the risk of LOAD in our primary analysis. Our data indicate that the association of the 16 previously reported SNPs by Grupe et al. (2007), if it exists, is not statistically significant in our study population. Furthermore, the additional six SNPs in positional and biological candidate regions for LOAD also showed no association. With 80% power at $\alpha = 0.05$ for 1,000 cases and 1,000 controls our sample size had sufficient power to detect small differences. For 12 SNPs, power was sufficient to detect an odds ratio (OR) of 1.2 or lower. For the other 10 SNPs, power was sufficient to detect ORs between 1.22 and 1.42. Ten of the 16 significant SNPs reported by Grupe et al. (2007) had ORs between 1.2 and 1.4 and they would have been detected in our sample. However, the ORs for the other 6 reported significant SNPs were only 1.07–1.12, and they could have been missed in our sample.

Genome-wide association studies allow detection of new genes not previously implicated based on the known biology of the disease. These studies may shed light on novel genes and SNPs that affect risk for complex diseases. Grupe et al. (2007) examined the whole genome for association of putative functional variants, instead of focusing only on the genes that are known biological and/or positional candidates. Many possible associations have been found without the immediate knowledge of their biological link to the disease. This information may help us understand the disease biology in greater detail in the future (Drazen et al., 2007). There are numerous difficulties when utilizing genome-wide association studies to identify candidate genes for complex diseases (Baron, 2001). It may be difficult to differentiate between genes with true association and those that are more common due to ethnic variation. Our study was ethnically homogenous of Caucasian descent, and from a single geographical location. Although the population from Grupe et al. (2007) was also of Caucasian descent, they were derived from five independent case-control cohorts from different locations. Since the category "Caucasian" is broad and individuals may still vary widely in genetic history, significant genetic variation may be present. One explanation why we were unable to replicate the previously reported findings is that the reported significant SNPs were not functional but rather in LD with nearby casual SNPs and this LD was not present in our sample. Another possible explanation is that since these SNPs had modest effect sizes, they were not amenable in all samples due to either population heterogeneity or discrepancies in AD diagnosis. The latter is likely not a reason; our Alzheimer's Disease Research Center has a very high confirmation rate (> 95%) between clinical diagnosis of AD cases and autopsy confirmation. Although only 8% of our AD cases were autopsy-confirmed, we found no statistically significant difference for any SNP between clinically diagnosed and autopsy-confirmed AD cases.

Different strategies may be used to replicate the data from a genome-wide study. An exact replication strategy attempts to replicate the results from a past study exactly, looking specifically at the SNPs identified as possible risk alleles. This type of strategy is balanced in terms of cost and the ability to replicate a maximum number of loci (Clarke et al., 2007). Our study design was an exact strategy. With such a focus we may have missed any other SNPs in these genes with an association with the risk of AD. Another possible strategy includes the exact replication of risk SNPs as well as surrounding SNPs in the same gene; this has been termed a local replication strategy. This study design is more costly but may have the ability to pick up associations missed by the exact strategy. If there are other SNPs in the gene of interest in LD with a casual allele this strategy may have the ability to find them. However, this approach can also reveal different loci than those initially identified, making interpretation difficult and usually requiring another validation study (Clarke et al., 2007).

Although no significant association was observed with the disease risk, one SNP revealed association with AAO and two SNPs showed association with disease duration. The association of the less favorable quantitative phenotypes (i.e., lower AAO and shorter disease duration) was with the same alleles that showed association with AD risk in the original study of Grupe et al. (2007). The C allele of SNP 10, *MYH13*/rs2074877 (CC + CT genotypes) was associated with lower AAO than the TT genotype (72.57 ± 6.00 vs. 74.55 ± 6.27 ; $p = 0.0004$). In addition to disease risk genetics also affects AAO with estimated heritability of > 40% (Li et al., 2002). A simulation study has suggested that in addition to *APOE*, seven other genes may be involved in affecting the variation in AAO (Daw et al., 2000) and the *MYH13* gene could be one of the sought genes. *MYH13* is located on chromosome 17p13 and codes for myosin heavy chain 13 that is expressed primarily in extrinsic eye muscle and thus its role in AD is uncertain. It is also not a positional candidate gene for AD risk or AAO of AD. Two SNPs located on chromosome 1q21 (*CTSS*/rs41271951 and *FAM63A*/rs41310885) showed significant association with disease duration where the less common allele at each SNP (G and A, respectively) was associated with shorter disease duration before death ($p = 0.006$, $p = 0.0014$, respectively). The two SNPs were in LD ($D' = 0.908$; $r^2 = 0.693$) and thus appear to represent

one association. Interestingly, the *CTSS* SNP alters an amino acid in cathepsin S, the levels of which are increased in AD brains (Lemere et al., 1995) and it also appears to regulate the generation of A β -peptide (Munger et al., 1995). To our knowledge these associations have not been reported previously and thus they should be treated as provisional findings until confirmed in independent samples. It is also possible that these significant associations are due to chance because multiple comparisons were performed, and in the case of disease duration analysis data were available from only 80 AD cases.

Future studies are needed in order to examine these genes with potential biological or positional significance, in greater detail. This may be done by investigating all common tag SNPs and rare variants present in these genes. However, even when all the SNPs in the genes have been studied, confounding factors such as the make up of the population and environment may still act to affect the outcome of the disease.

Acknowledgments

This study was supported by the National Institute on Aging (NIA) grants AG13672 and AG05133. We thank Dr. Eleanor Feingold and Dr. Candace M. Kammerer for their guidance in the statistical analysis.

References

- Baron M. The search for complex disease genes: fault by linkage or fault by association? *Mol Psychiatry* 2001;6:143–149. [PubMed: 11317215]
- Clarke GM, Carter KW, Palmer LJ, Morris AP, Cardon LR. Fine mapping versus replication in whole-genome association studies. *Am J Hum Genet* 2007;81:995–1005. [PubMed: 17924341]
- Daw EW, Payami H, Nemens EJ, Nochlin D, Bird TD, Schellenberg GD, Wijsman EM. The number of trait loci in late-onset Alzheimer disease. *Am J Hum Genet* 2000;66:196–204. [PubMed: 10631151]
- Drazen JM, Phimister EG. Publishing genomewide association studies. *N Engl J Med* 2007;357:496.
- Dupont WD, Plummer WD Jr. Power and sample size calculations for studies involving linear regression. *Control Clin Trials* 1998;19:589–601. [PubMed: 9875838]
- Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 2006;63:168–174. [PubMed: 16461860]
- Grupe A, Abraham R, Li Y, Rowland C, Hollingworth P, Morgan A, Jehu L, Segurado R, Stone D, Schadt E, Karnoub M, Nowotny P, Tacey K, Catanese J, Sninsky J, Frayne C, Rubinsztein D, Gill M, Lawlor B, Lovestone S, Holmans P, O'Donovan M, Morris JC, Thal L, Goate A, Owen MJ, Williams J. Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. *Hum Mol Genet* 2007;16:865–873. [PubMed: 17317784]
- Kamboh MI. Molecular genetics of late-onset Alzheimer's disease. *Ann Hum Genet* 2004;68:381–404. [PubMed: 15225164]
- Lemere CA, Munger JS, Shi GP, Natkin L, Haass C, Chapman HA, Selkoe DJ. The lysosomal cysteine protease, cathepsin S, is increased in Alzheimer's disease and Down syndrome brain: an immunocytochemical study. *Am J Pathol* 1995;146:848–860. [PubMed: 7717452]
- Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, Hubble JP, Koller WC, Pahwa R, Stern MB, Hiner BC, Jankovic J, Allen FA Jr, Goetz CG, Mastaglia F, Stajich JM, Gibson RA, Middleton LT, Saunders AM, Scott BL, Small GW, Nicodemus KK, Reed AD, Schmechel DE, Welsh-Bohmer KA, Conneally PM, Roses AD, Gilbert JR, Vance JM, Haines JL, Pericak-Vance MA. Age at onset in two common neurodegenerative diseases is genetically controlled. *Am J Hum Genet* 2002;70:985–993. [PubMed: 11875758]
- Munger JS, Haass C, Lemere CA, Shi GP, Wong WS, Teplow DB, Selkoe DJ, Chapman HA. Lysosomal processing of amyloid precursor protein to A β peptides: a distinct role for cathepsin S. *Biochem J* 1995;311:299–305. [PubMed: 7575468]

- R Development Core Team. R: A language and environment for statistical computing, version 2.2.0. R Foundation for Statistical Computing; Vienna, Austria: 2005. www.r-project.org
- Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, König IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007;357:443–453. [PubMed: 17634449]

Genotype and allele frequencies of 22 SNPs cases and controls

SNP Number*	Ref. SNP	Chromosome	Gene	Type	AD Status	n	Genotype Frequencies			p	Allele Frequencies		p
							CC	CG	GG		C	G	
1	rs3745833	19q13.42	GALP	missense	cases	1008	0.423	0.462	0.115	0.424	0.654	0.346	0.198
							0.399	0.471	0.130		0.634	0.366	
							AA	AG	GG		A	G	
2	rs8192708	20q13.31	PCK1	missense	cases	1009	0.800	0.138	0.115	0.057	0.869	0.131	0.015
							0.756	0.172	0.130		0.842	0.158	
							AA	AT	TT		A	T	
3	rs1554948	17p13.1	TNKI	silent	cases	1009	0.185	0.500	0.314	0.101	0.436	0.564	0.115
							0.224	0.473	0.303		0.460	0.540	
							AA	AT	TT		A	T	
4	rs11622883	14 q32.13	near SERPINA13	intergenic	cases	1007	0.214	0.495	0.291	0.931	0.462	0.538	0.955
							0.211	0.503	0.286		0.463	0.537	
							AA	AG	GG		A	G	
5	rs3800324	6p22.1	PGBDI	missense	cases	1009	0.000	0.083	0.917	0.411	0.042	0.958	0.974
							0.003	0.077	0.920		0.041	0.959	
							CC	CT	TT		C	T	
6	rs505058	1q22	LMNA	silent	cases	1009	0.185	0.500	0.314	0.101	0.436	0.564	0.115
							0.224	0.473	0.303		0.460	0.540	
							CC	CG	GG		C	G	
7	rs444013	6p22.1	UBD	3'UTR	cases	1009	0.253	0.520	0.227	0.049	0.513	0.487	0.168
							0.298	0.473	0.229		0.535	0.465	
							CC	CG	GG		C	G	

SNP Number*	Ref. SNP	Chromosome	Gene	Type	AD Status	n	Genotype Frequencies			p	Allele Frequencies		p
8	rs1859849	7p15.2	10 kb from <i>UBA52PI</i>	intergenic	cases	1009	0.058	0.386	0.556	0.276	0.251	0.749	0.491
							0.076	0.370	0.554		0.261	0.739	
9	rs9608099	22q11.23	<i>BCR</i>	5'UTR	cases	1005	0.096	0.457	0.448	0.354	0.324	0.676	0.319
							0.096	0.426	0.478		0.309	0.691	
10	rs2074877	17q13.1	<i>MYH13</i>	missense	cases	1009	0.214	0.495	0.291	0.931	0.462	0.539	0.955
							0.211	0.503	0.286		0.463	0.537	
11	rs11016976	10q26.3	<i>EBF3</i>	TFBS intron	cases	1009	0.024	0.250	0.726	0.354	0.149	0.851	0.192
							0.023	0.223	0.754		0.134	0.866	
12	rs2882676	15q26.1	<i>AGC1</i>	missense	cases	1007	0.144	0.452	0.404	0.674	0.370	0.630	0.476
							0.130	0.457	0.412		0.359	0.641	
13	rs41271951 (hCV15746640) [†]	1q21	<i>CTSS</i>	missense	cases	1009	0.851	0.141	0.008	0.121	0.922	0.078	0.924
							0.844	0.154	0.002		0.921	0.079	
14	rs41310885 (hCV22274641) [†]	1q21.2	<i>FAM63A</i>	intron	cases	1008	0.006	0.116	0.878	0.059	0.064	0.936	0.924
							0.001	0.138	0.861		0.070	0.930	
					controls	1007	CC	CT	TT		C	T	

SNP Number*	Ref. SNP	Chromosome	Gene	Type	AD Status	n	Genotype Frequencies			p	Allele Frequencies		p
15	rs13022344	2q33.1	TRAK2	missense	cases	1009	0.107	0.453	0.440	0.912	0.333	0.667	0.797
						controls	0.108	0.444	0.449		0.330	0.670	
							AA	AG	GG		A	G	
16	rs6907175	6q24.1	LOC651924	3'UTR	cases	967	0.257	0.450	0.275	0.061	0.500	0.500	0.542
						controls	0.259	0.502	0.239		0.510	0.490	
							CC	CT	TT		C	T	
17	rs501120	10q11.21	100 kb from CXCL12	intergenic	cases	1008	0.019	0.258	0.723	0.899	0.148	0.852	0.688
						controls	0.019	0.249	0.732		0.143	0.857	
							AA	AC	CC		A	C	
18	rs2943634	2q36.3	532 kb from IRS1	intergenic	cases	1000	0.127	0.431	0.442	0.793	0.343	0.658	0.497
						controls	0.118	0.428	0.454		0.332	0.668	
							CC	CT	TT		C	T	
19	rs17228212	15q22.33	SMAD3	intron	cases	1009	0.075	0.373	0.552	0.945	0.262	0.738	0.954
						controls	0.079	0.368	0.554		0.262	0.738	
							CC	CG	GG		C	G	
20	rs1333049	9p21.3	LOC401495		cases	1009	0.249	0.488	0.264	0.201	0.493	0.507	0.704
						controls	0.223	0.526	0.250		0.487	0.513	
							AA	AG	GG		A	G	
21	rs6922269	6q25.1	MTHRD1L	intron	cases	1008	0.074	0.392	0.534	0.704	0.270	0.729	0.414
						controls	0.067	0.384	0.549		0.259	0.742	
							CC	CT	TT		C	T	
22	rs7901695	10 q25.3	TCF7L2	intron	cases	1009	0.094	0.419	0.487	0.282	0.304	0.696	0.125
						controls	0.113	0.426	0.461		0.326	0.674	