ELSEVIER

Journal of the Neurological Sciences 236 (2005) 17 – 24

Journal of the Neurological Sciences

www.elsevier.com/locate/jns

Chromosome 21 *BACE2* haplotype associates with Alzheimer's disease: A two-stage study

Liisa Myllykangas [a], Fabienne Wavrant-De Vrièze [b], Tuomo Polvikoski [a,c], Irma-Leena Notkola [d], Raimo Sulkava [e], Leena Niinistö [f], Steven D. Edland [g], Sampath Arepalli [b], Omanma Adighibe [b], Danielle Compton [b], John Hardy [b], Matti Haltia [a], Pentti J. Tienari [h,j,*]

[a]*Departments of Pathology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland*
[b]*Laboratory of Neurogenetics, NIA/NIH, Bethesda, MD, USA*
[c]*Department of Neuropathology and Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne, UK*
[d]*Department of Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland*
[e]*Division of Geriatrics, Department of Public Health and General Practice and Medicine, University of Kuopio, Kuopio, Finland*
[f]*Katriina Geriatric Hospital, Vantaa, Finland*
[g]*Mayo Clinic Rochester, Rochester, MN, USA*
[h]*Central Military Hospital, Helsinki, Finland*
[j]*Department of Neurology and Neuroscience Programme, Biomedicum-Helsinki, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland*

Received 8 July 2004; received in revised form 3 January 2005; accepted 18 April 2005
Available online 14 July 2005

Abstract

Genetic linkage studies have provided evidence for a late-onset Alzheimer's disease (AD) susceptibility locus on chromosome 21q. We have tested, in a two-stage association study, whether allelic or haplotype variation of the beta-amyloid cleaving enzyme-2 (*BACE2*) locus on chromosome 21q affects the risk of late-onset AD. In stage-1, an unselected population-based sample of Finns aged 85 years or over ($n=515$) was analysed. Neuropathologic examination including beta-amyloid load quantification was possible in over 50% ($n=264$) of these subjects. AD patients ($n=100$) and controls ($n=48$) were defined by modified neuropathological NIA-RI criteria. Positive associations were taken as a hypothesis, and tested in stage-2 using 483 AD families from the USA. Four single nucleotide polymorphisms (SNPs) of BACE2 gene were tested in stage-1. A SNP close to exon-6 was associated with neuropathologically verified AD ($p=0.02$) and also with beta-amyloid load in non-selected autopsied subjects after conditioning with *APOE* genotype ($p=0.001$). In haplotype analysis a specific, relatively common haplotype (H5) was found to associate with AD ($p=0.004$) and a second haplotype (H7) showed a weaker association with protection against AD ($p=0.04$). In stage-2, the SNP association was not replicated, whereas the haplotype H5 association was replicated ($p=0.004$) and a trend to association was found with the putative protective haplotype H7 (two-sided $p=0.08$). *BACE2* haplotype association with AD in two independent datasets provides further evidence for an AD susceptibility locus on chromosome 21q within or close to *BACE2*.

Keywords: Alzheimer; BACE2; Secretase; Chromosome 21; Amyloid; Neuropathology

1. Introduction

Alzheimer's disease, the most common cause of dementia in the elderly, is characterized clinically by progressive loss of cognitive functions, and neuropathologically by accumulation of neuritic plaques and neurofibrillary tangles

* Corresponding author. Neuroscience Programme, Biomedicum-Helsinki, University of Helsinki, POB 700, FI-00290 Helsinki, Finland. Tel.: +358 9 47171960; fax: +358 9 47171964.
E-mail address: pentti.tienari@hus.fi (P.J. Tienari).

0022-510X/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jns.2005.04.008

in the brain. Definite diagnosis of AD requires neuropathological examination. According to hospital-based studies, approximately 80% of the subjects clinically diagnosed as AD have also neuropathologically verified AD [1,2], but in the population-based studies concerning very elderly subjects the clinical diagnostics may be more problematic [3].

Although epidemiological studies have clearly shown that there is a strong genetic component underlying the common late-onset form of the disease, only one universally accepted genetic risk factor, the ε4-allele of apolipoprotein E (*APOE*) has been identified. As *APOE* ε4 is neither sufficient nor necessary to cause the disease, it is likely that other genetic risk factors exist. Numerous AD loci have been suggested by previous studies, but only few of these findings have been replicated [4]. One of the most promising loci, on the basis of several linkage studies, is on chromosome 21q [5–9].

Mismetabolism of the amyloid precursor protein (APP), particularly increased generation of beta-amyloid (Aβ) peptide, is believed to play a crucial role in the pathogenesis of AD [10]. The generation of Aβ requires cleavage of APP by β- and γ-secretases. A third secretase activity, termed α-secretase, cleaves within the Aβ-sequence thus precluding the generation of full length Aβ. The newly identified beta-amyloid cleaving enzyme-2 (BACE2), a homologue of the major β-secretase BACE1, has been demonstrated to have a minor cleavage site at the β-site of APP and a major cleavage site within the Aβ-region close to the α-secretase site [11]. Both *APP* and *BACE2* are among the candidate genes of AD located on chromosome 21q.

The aim of this study was to test, using genetic association analysis, whether *BACE2* on chromosome 21q would harbor a susceptibility locus for late-onset AD. Since many initial genetic associations with AD have not been replicated in other datasets or populations we used a two-stage approach. In stage-1, four single nucleotide polymorphisms within *BACE2* were analyzed in a population-based and neuropathologically verified dataset of very elderly Finns. Initial associations found in stage-1 were taken as hypotheses that were tested in stage-2 in a dataset consisting of AD families from the USA. Previous studies of these AD families have suggested linkage with chromosome 21q [7]. Our results provide evidence, in both datasets, for a *BACE2* haplotype association with AD.

2. Subjects and methods

2.1. Subjects

2.1.1. Dataset-1: the Vantaa-85+ population

The Vantaa-85+ study population has proved useful in genetic studies of late-onset AD [12–14] and was therefore chosen as the hypothesis setting dataset-1. The Vantaa-85+ study population includes all persons aged 85 years or over who were living in the city of Vantaa (Southern Finland), on April 1, 1991. Fig. 1 illustrates the overall structure of the study population. Of the 601 eligible subjects, *BACE2* genotyping was possible in 515 study subjects, and of these 264 had been neuropathologically examined. Using the modified NIA-RI criteria, there were 100 AD patients and 48 controls. 84% of the AD patients and 81% of the controls were females. The mean age at death of the AD patients was 92.3 years, and 91.7 years for the controls. Mean age at onset of the AD patients was 86.1 years. All specimen collections were performed after obtaining informed consent from the subjects, or their relatives or legal guardians, with the approval of the Ethical Review Committee of the Health Center of the City of Vantaa and in compliance with the Declaration of Helsinki.

2.1.2. Dataset-2: AD families from the USA

The US sibpair series consisted of 483 nuclear AD families ascertained by the NIMH-AD Genetics Initiative and obtained from the NIMH (1007 samples) and the Indiana Alzheimer's Disease Center National Cell Repository (NIA, 365 samples). The families included sibpairs affected by probable or confirmed AD and the oldest unaffected relatives. Total number of samples was 1372. There were 300 unaffected siblings (59% female) with a mean age of 74.8 years and 1072 AD cases (73% female) with a mean age of 81.5 years and a mean age at onset of 73.3 years. The percentage of female patients was lower and the mean age of onset was 12.8 years younger than in the Vantaa-85+ population. The U.S. sibpair families studied here partially overlap with the samples used in previous studies. Of the 483 families in the present study 172 (35.6%) were included in the study by Myers et al. [15].

2.2. Neuropathological examination

The diagnosis of neuropathological AD was established according to the modified NIA-RI criteria [3]. Briefly,

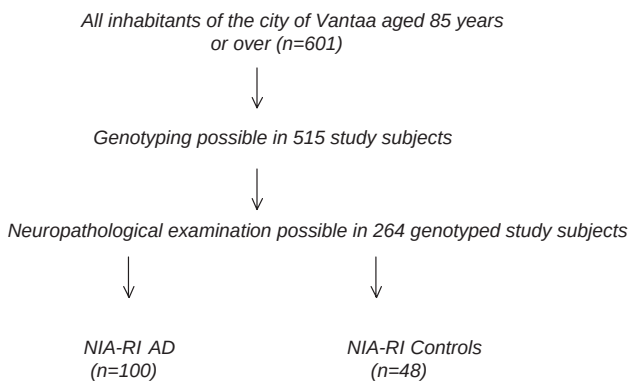

Fig. 1. The Vantaa-85+ study population.

diagnosis of AD according to the modified NIA-RI criteria required dementia and plaque scores "moderate" or "frequent" according to the CERAD protocol [16], and stages IV–VI of neurofibrillary pathology [17]. The neuropathological controls were defined as non-demented individuals with no neuritic plaques or with the CERAD neuritic plaque score "sparse". In addition, the controls had to have no more than stage II of neurofibrillary pathology. The neocortical Aβ load was quantified as the percentage of cerebral cortex covered by methenamine silver-positive plaques, as described earlier [12].

2.3. Genotyping

DNA was extracted from peripheral blood or tissue specimens according to standard procedures. Two single nucleotide polymorphisms (SNP) were obtained from the public SNP database (http://www.ncbi.nlm.nih.gov/SNP/). These SNPs were rs1467756 (here termed intron-2A) located on *BACE2* intron 2, at position 28,205,646 of the contig NT_011512, and rs766850 (here termed intron-2B) located on BACE2 intron 2 at position 28,207,213 of NT_011512. These SNPs were genotyped according to standard PCR and RFLP methods. The PCR primers were 5′-ACCGGAGT-TACGTCTTCAGC-3′ and 5′-TTCCTTCTCCCACCCT AAGC-3′ (intron 2A) and 5′-AGTCCCTTGCTGTGTGT-CAT-3′ and 5′-CAGGAAAGTCACCCA GAAGA-3′ (intron 2B). The PCR products were digested with restiction enzymes *Cla*I and *Msp*I, respectively. *BACE2* exon 6 SNP (rs2252576) located 10 bases upstream of exon 6 at position 28,275,689 of NT_011512 and exon 8 SNP (rs1046210, a synonymous change at Asp364) at position 28,283,182 of NT_011512 were genotyped as described earlier [18]. These SNPs were termed in the earlier publication as exon-5 and exon-7 [18], before the discovery of an additional exon in the 5′ end of the gene.

2.4. Statistical analysis

In single marker analyses of the Vantaa-85+ study (stage-1) alleles and genotypes were counted and their distribution between groups was compared by the χ^2 or Fisher's exact test. Hardy–Weinberg equilibrium of the markers was analyzed using the HWSIM software (http://krunch.med.yale.edu/hwsim/). Linkage disequilibrium between pairs of markers was analyzed by the LINKD program [19] (http://krunch.med.yale.edu/haplo), and expressed as Lewinson's linkage disequilibrium coefficient (D') and p-value. The expectation–maximization (EM) algorithm, as implemented in the HAPLO software, was used to estimate haplotype frequencies [20,21] (http://krunch.med.yale.edu/haplo). The difference in estimated haplotype frequencies in AD patients vs. controls was compared by using the formula $2\ln L_{AD} + 2\ln L_{Controls} - 2\ln L_{Overall} \sim \chi^2$. Given the sample size, we a priori decided to analyze *BACE2* haplotypes using three markers at a time. *APOE* ε4 stratification was not applied to the haplotype analysis in the Vantaa-85+ dataset since the number of subjects would have been too low for meaningful analysis. Based on simulation studies the accuracy of EM-algorithm decreases with such small sample size [22]. Stage-1 served for hypothesis setting, corrections for multiple comparisons were not done. The p-values represent uncorrected nominal p-values. The number of independent tests is very difficult to define, when analyzing tightly linked markers exhibiting linkage disequilibrium with each other. In stage-1 the number of markers was 4 and the theoretical number of haplotypes composed by these markers is 16. On the basis of these definitions the number of tests was at least 20 in the explorative stage-1 (4 individual markers + 16 haplotypes). In stage-2 the number of tests was 10 (3 SNP tests + 7 haplotype tests, including pooled rare haplotypes).

Family association analysis was performed in stage-2 using the Family Based Association Tests (FBAT) software [23] in which the genotype information of unaffected sibs is used, when parents are missing. To analyze the haplotype transmission, we used the same software, which is able to deal with multilocus haplotypes, even in the presence of phase uncertainty and missing parental genotypes. Haplotypes with a frequency of less than 5% were pooled. Multiple nuclear families were included from extended pedigrees. The empirical variance calculation of FBAT was used for the estimation of p-values [24]. Our primary goal was to analyze allelic/haplotypic association between AD and *BACE2*. We tested the default null hypothesis of FBAT: no linkage and no association between the marker influencing AD risk and *BACE2* gene. In the US sibpair series (dataset-2) we also performed non-parametric linkage analysis with the *BACE2* markers using the SPLINK 1.09 software [25].

3. Results

3.1. BACE2 single marker analysis in the Vantaa-85+ population (stage-1)

We analyzed four SNPs within the *BACE2* gene in the neuropathologically verified AD cases vs. controls. Two of the SNPs were located in intron-2, one was located just upstream of exon-6, and one was a silent substitution in exon-8. An association was found between the BACE2 exon-6 polymorphism and AD (allele-wise $p = 0.02$) (Table 1). The odds ratio for AD was 0.42 (95% CI 0.21–0.86, T/T + C/T vs. C/C) for subjects with the exon-6 T-allele and 1.72 (95% CI 0.44–6.75, C/C + C/T vs. T/T) for subjects with exon-6 C-allele, suggesting a protective effect associated with the T-allele.

Stratification of the subjects according to the carrier status of *APOE* ε4-allele showed that the exon-6 association largely arose from the *APOE* ε4-negative subpopulation

Table 1
BACE2 genotypes and alleles in the neuropathologically verified AD patients and controls (Vantaa-85+ population)

	AD *n* (%)	Controls *n* (%)	χ^2
Intron-2A			
C/C	14 (14.0%)	2 (4.4%)	2.90
C/T	44 (44.0%)	30 (62.5%)	2.22
T/T	42 (42.0%)	16 (33.3%)	0.62 (χ^2=5.72, 2 *df*, p=0.06)
C	72 (36.0%)	34 (35.4%)	0.00
T	128 (64.0%)	62 (64.6%)	0.01 (χ^2=0.01, 1 *df*, p=0.92)
Intron-2B			
T/T	2 (2.0%)	1 (2.1%)	n.a.
T/C	32 (32.3%)	12 (25.0%)	0.53
C/C	65 (65.7%)	35 (72.9%)	0.25 (χ^2=0.78, 1 *df*, p=0.38)
T	36 (18.2%)	14 (14.6%)	0.49
C	162 (81.8%)	82 (85.4%)	0.10 (χ^2=0.59, 1 *df*, p=0.44)
Exon-6			
C/C	72 (72.0%)	25 (52.1%)	1.97
C/T	23 (23.0%)	19 (39.6%)	3.14
T/T	5 (5.0%)	4 (8.3%)	0.59 (χ^2=5.70, 2 *df*, p=0.06)
C	167 (83.5%)	69 (71.9%)	1.10
T	33 (16.5%)	27 (28.1%)	4.32 (χ^2=5.42, 1 *df*, p=0.02)
Exon-8			
C/C	34 (35.1%)	15 (31.3%)	0.14
C/T	48 (49.5%)	24 (50.0%)	0.00
T/T	15 (15.5%)	9 (18.8%)	0.21 (χ^2=0.35, 2 *df*, p=0.85)
C	116 (59.8%)	54 (56.3%)	0.14
T	78 (40.2%)	42 (43.8%)	0.19 (χ^2=0.33, 1 *df*, p=0.56)

Chi-square test was used to determine the *p*-values. Due to the low number of observations the intron-2B T/T genotype was pooled with T/C. The odds ratio for AD was 0.42 (95% CI 0.21–0.86) for subjects with the exon-6 T-allele (T/T+T/C) as compared to those without it (C/C). The odds ratio for AD was 1.72 (95% CI 0.44–6.75) for subjects with exon-6 C allele (C/C+C/T) as compared to those without it (TT). n.a.: not applicable.

(p=0.01). There were very few ε4-positives in the controls, hence it was not possible to perform statistical testing on this subgroup.

We next analysed whether the association with the exon-6 SNP could be detected using the extent of neocortical Aβ deposition as a quantitative variable. This analysis was done in the whole neuropathologically examined sample (n=264), which consisted of the 100 AD cases, 48 controls and those 116 subjects, who did not fulfill the neuropathological criteria required for inclusion in the AD nor control group. Statistical analysis was performed by comparing the same two genotype groups as above: T/T+C/T vs. C/C. In the whole sample, there was no significant association between the exon-6 SNP and the cortical Aβ load (p=0.12). In the *APOE* ε4 negative subgroup an association was found (p=0.005, data not shown). Because also *APOE* ε2 bears an impact on Aβ-load, in addition to ε4 [12], we made a further analysis of Aβ-load in the ε3/ε3 subgroup. This is a more standardized group than the *APOE* ε4-negatives and excludes the impact of both ε2 and ε4 [14]. In the ε3/ε3 subgroup an even stronger association between *BACE2*=exon-6 SNP and Aβ load was found (p=0.001) (Fig. 2).

3.2. BACE2 haplotype analysis in the Vantaa-85+ population (stage-1)

In the total unselected Vantaa-85+ population (n=515) all *BACE2* SNPs were in Hardy–Weinberg equilibrium (HWE) (data not shown). In the neuropathologically verified AD patients all markers were in HWE. In the neuropathologically verified controls one of the markers (intron-2A) deviated (p<0.05) from the HWE due to a higher number than expected of heterozygous subjects (see Table 1).

Linkage disequilibrium was observed at the level of p<0.001 between marker pairs intron-2A/intron-2B (D' 0.21), intron-2B/exon-6 (D' 0.73), intron-2B/exon-8 (D' 0.37), exon-6/exon-8 (D' 0.80), and at the level of p<0.01 between markers intron-2A/exon-8 (D' 0.17). No significant linkage disequilibrium was found between markers intron-2A/exon-6 (D' 0.08).

BACE2 haplotype associations with AD were analyzed using three-site haplotypes (haplotypes consisting of three SNPs), and the four possible combinations of markers were tested. Haplotype frequencies were estimated for AD patients and controls by applying the expectation–maximization (EM) algorithm [20,21]. The global haplotype distributions in AD and controls were compared first. Two out of the four *BACE2* marker combinations showed significant differences in the haplotype distributions in patients vs. controls with

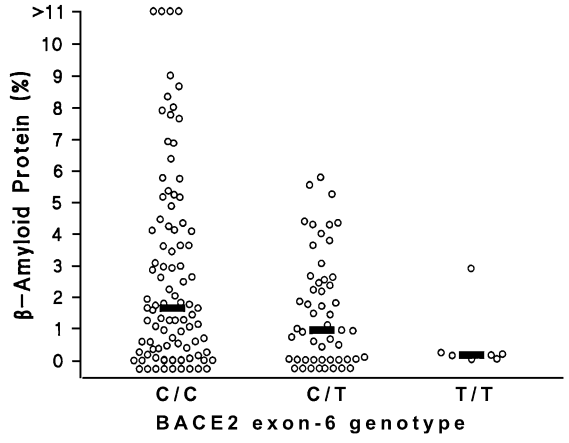

Fig. 2. Beta-amyloid load in 155 APOE3/3 individuals according to the BACE2 exon-6 genotype. Dots represent individual values for percentage of neocortex covered by beta-amyloid protein (bars show medians for each *BACE2* genotype). Median percentage was 1.64 in C/C subjects (n=95) vs. 0.80 in C/T+T/T subjects (n=60), p=0.001 (t-test).

Table 2
BACE2 haplotype frequency estimates in neuropathologically verified AD and controls

	AD (194 Chromosomes) Frequency (S.E.)	Controls (96 Chromosomes) Frequency (S.E.)
A. Intron-2A/exon-6/exon-8		
H1 (CCC)	0.08 (0.024)	0.14 (0.056)
H2 (CCT)	0.23 (0.032)	0.15 (0.048)
H3 (CTC)	0.05 (0.019)	0.05 (0.034)
H4 (CTT)	0.00 (0.000)	0.01 (0.013)
H5 (TCC)*	0.35 (0.037)	0.15 (0.059)
H6 (TCT)	0.18 (0.027)	0.27 (0.053)
H7 (TTC)	0.11 (0.026)	0.22 (0.047)
H8 (TTT)	0.00 (0.000)	0.00 (0.000)
		$\chi^2=14.26$, 5 *df*, global $p=0.01$
B. Intron-2B/exon-6/exon-8		
CCC	0.10 (0.037)	0.17 (0.049)
CCT	0.15 (0.046)	0.07 (0.035)
CTC	0.04 (0.033)	0.00 (0.000)
CTT	0.07 (0.040)	0.11 (0.032)
TCC	0.42 (0.038)	0.39 (0.055)
TCT	0.15 (0.030)	0.22 (0.042)
TTC	0.04 (0.019)	0.00 (0.000)
TTT	0.03 (0.021)	0.03 (0.022)
		$\chi^2=14.20$, 4 *df*, global $p=0.007$
C. Other marker combinations		
Intron-2A/intron-2B/exon-6		$\chi^2=6.83$, 5 *df*, global $p=0.23$
Intron-2A/intron-2B/exon-8		$\chi^2=12.52$, 6 *df*, global $p=0.05$

Global p-values for the haplotype distributions are given for each marker combination. Global p-values were calculated using the formula $2\ln L_{AD}+2\ln L_{Controls}-2\ln L_{Overall}\sim\chi^2$. Haplotypes with overall frequencies less than 0.05 were excluded, when calculating the global p-values. S.E.: standard error (Jackknife).

* $p=0.004$ for haplotype H5 (Z-test).

global p-values of 0.007 and 0.01 (Table 2A and B). A third marker combination was of borderline significance (global p-value 0.051), while the fourth was clearly non-significant (Table 2B, all haplotype frequencies are available upon request).

Interestingly, using the marker combination intron-2A/exon-6/exon-8 there was a specific, relatively common haplotype (H5), which associated with AD (Table 2). The estimated frequency of the H5 haplotype was 0.35 in AD vs. 0.15 in controls ($p=0.004$, Z-test) (Table 2A). Conversely, the frequency of the H7 haplotype (Table 2A) was decreased in AD vs. controls (0.11 vs. 0.22, $p=0.04$, Z-test).

Intron-2A/exon-6/exon-8 haplotype was selected for further testing in stage-2 because this combination of markers defined a relatively common AD-associated haplotype variant (H5). H5 haplotype was also the one, which showed most striking difference in frequency when comparing the patients and the controls. Furthermore, we reasoned that a common variant would be more likely to be found in another dataset than a rare variant.

3.3. Hypothesis testing in AD families (stage-2)

The initial associations detected with exon-6 and the intron-2A/exon-6/exon-8 haplotypes were taken as a hypothesis that was tested in stage-2 in the dataset, which consisted of 483 AD families from the USA. All markers, except exon-6 ($p<0.001$) were in HWE in controls. The HWE deviation of exon-6 in the controls was mainly contributed by the increased frequency of T/T homozygotes (not shown). No deviations from HWE were found in the AD group (not shown).

In the family based association analysis none of the SNPs were associated with AD (Table 3). In haplotype association analysis the association with the haplotype H5 (TCC) was replicated ($p=0.004$, Table 3), and a trend to decreased frequency of the haplotype H7 (TTC) was also found (two-sided $p=0.08$, Table 3). The majority of the families were *APOE* ε4 positives (all cases had ε4) and the findings with the H5 and H7 haplotypes were mainly contributed by these (H5: $p=0.002$, haplotype found in 103 families; H7: $p=0.06$, haplotype found in 69 families). The global p-value in the *APOE* ε4 positives was 0.001. The number of *APOE* ε4 negative families was much smaller, these families did not show any association with H5 or H7 when analyzed alone (H5: $p=0.79$, haplotype found 37 families; H7: $p=0.79$, haplotype found in 23 families).

We also performed non-parametric linkage analysis using the SPLINK software [25]. We did not find statistically significant evidence for linkage between AD and *BACE2* either using single SNPs or haplotypes as markers (all p-values >0.1, data not shown), This is consistent with a gene conferring a small effect on disease risk, detectable by means of association rather than linkage.

Table 3
Association analysis of the *BACE2* SNPs and haplotypes in AD families from the USA (stage-2)

Marker	Frequency[a]	No. of families	p-value
Intron-2A T/C	0.71/0.29	113	0.68
Exon-6 T/C	0.27/0.73	104	0.38
Exon-8 T/C	0.41/0.59	114	0.09
H1 (CCC)	0.11	79	0.63
H2 (CCT)	0.10	69	0.90
H3 (CTC)	0.06	42	0.71
H4 (CTT)	0.01	3	0.27
H5 (TCC)	0.27	132	0.004
H6 (TCT)	0.27	138	0.31
H7 (TTC)	0.16	95	0.08
H8 (TTT)	0.02	16	0.57

The Monte Carlo permutation test was used. Global $p=0.08$.

[a] Estimated allele/haplotype frequencies in the parental chromosomes.

4. Discussion

This study reports findings of a two-stage study testing whether allelic or haplotype variation in the *BACE2* locus affects the risk of late-onset AD. In stage-1, a SNP located upstream of *BACE2* exon 6 showed an association with AD ($p=0.02$), and also associated with the extent of Aβ deposition in a population-based sample of *APOE* ε3/ε3 subjects ($p=0.001$). Haplotype analysis revealed that the estimated haplotype frequencies in the *BACE2* region differ significantly in AD patients vs. controls. A specific, relatively common haplotype H5 was found to be associated with AD ($p=0.004$). In stage-2, using AD families from the USA, the H5 haplotype association was replicated ($p=0.004$), while the exon-6 SNP association was not confirmed. An additional trend to negative association was detected with haplotype H7 (2-sided *p*-values 0.04 in stage-1 and 0.08 in stage-2).

The *APOE* stratified results appear contradictory. In stage-1, most of the evidence for the exon-6 SNP association came from the *APOE*4 negative subgroup, while in stage-2 the haplotype association was mainly contributed by the *APOE*4-positive subgroup. However, in stage-1 there were very few *APOE*4-positive controls and in stage-2 only a low number of *APOE*4-negative patients/families. Hence, due to the sample sizes after stratification, we think that the present results do not allow us to make any conclusions on possible interaction with *APOE*.

Association studies in AD have been prone to produce positive results, which cannot be replicated in other datasets [4]. Case-control design may produce false positive associations due to population heterogeneity, i.e. cases and controls stem from distinct subpopulations with distinct genetic backgrounds. Genealogical studies have suggested that this may be the case even when the patients and controls are drawn from an unselected population sample [26,27]. Family-based studies are relatively free of population heterogeneity artifacts since both the case and control chromosomes are drawn from the same families and, hence, from the same ethnic background. For these reasons we used a two-stage approach with a hypothesis-setting dataset and a second family-based hypothesis-testing dataset.

The neuropathologically verified Vantaa-85+ dataset has special value in genetic studies due to the precision in the definition of AD and non-AD phenotypes. All patients of the Vantaa-85 + Study had neuropathologically verified AD, and a rigorous selection was applied to the controls with respect to beta-amyloid and neurofibrillary pathology. Therefore the AD cases and the controls can be considered as extremely discordant phenotypes drawn from the same unselected Vantaa-85+ population. The significance of neuropathological definition of both cases and controls is highlighted by our previous report showing that the odds ratio for AD of *APOE* ε4 is 19.6 in the neuropathologically verified Vantaa-85+ subpopulation [13], which is among the strongest reported to date. The corresponding odds ratio is only 2.88, when the clinical diagnostic criteria are used to differentiate patients and controls [13]. Our previous report of an association of AD with *alpha-2-macroglobulin* was confined to the neuropathologically defined group of AD and controls, no associations were detected in the clinically defined cases and controls [13]. We also tested the *BACE2* exon-6 SNP and H5 haplotype in the clinically defined subpopulation and observed no associations (data not shown). The poor correlation of genetic findings using clinical vs. neuropathological criteria is probably due to the problems in clinical definition of AD (and controls) in this very elderly population. In our recent analysis [3] more than half of the neuropathologic AD cases were found in other groups than clinical AD group. Moreover, one-third of the cases in the clinical AD group did not have neuropathologic AD and 21% of the non-demented subjects (controls) fulfilled the criteria for neuropathologic AD.

We were able to replicate the haplotype H5 association with AD (2-sided *p*-values 0.004 in stage-1 and 0.004 in stage-2). A trend towards a negative (or protective) haplotype association with H7 was also detected (2-sided *p*-values 0.04 and 0.08). The individual SNP analyses did not produce any replicable results here. However, haplotype-analyses tend to be more sensitive in detecting genetic associations than single SNP analyses since the population variation of a given locus is more efficiently revealed by haplotype analysis than by single marker analyses [28–30].

A potential concern in our haplotype analyses is the observed deviation from the HWE in the controls/unaffected siblings. Often markers exhibiting marked HWE deviation are excluded from association analysis. However, in the present study the HWE deviation was minor in the Vantaa-85+ sample ($0.01<p<0.05$), and haplotype association was found with AD, when the deviating marker (Intron-2A), was not included to the haplotype. Random population sampling is assumed in HWE, but here the controls/unaffected siblings were highly selected, which may also contribute to the observed HWE deviation. In the Vantaa-85+ sample the controls consisted of subjects with extremely little amyloid and neurofibrillary pathology in the brains, whereas in the AD families unaffected siblings were subjects who had remained free of AD symptoms despite multiply affected sibs. In previous studies deviations from HWE have been found with SNPs within or close to *APOE* in AD cases, indicating that selective sampling and analysis of a phenotypically relevant gene may result in deviations from HWE [30]. Genotyping error is often the cause of HWE deviation. We consider this possibility unlikely in our study since all genotyping in the sibpair series were done in duplicate to minimize errors.

Several candidate genes reside on chromosome 21q including *APP* and *BACE2*, which are located physically approximately 17 Mb, and in terms of recombination frequency more than 20 cM, apart from each other. Our data strongly argue that the putative susceptibility gene marked by the *BACE2* haplotype H5 would be closer to

BACE2 than *APP* since allelic/haplotypic associations are not expected to extend as long distances as >20 cM. In association studies the candidate region for the location of the susceptibility gene is expected to be at most a few megabases/centimorgans from the associated marker, and some investigators are in favor of much shorter distances [31]. Although *BACE2* is an obvious functional, and now also a positional candidate gene for AD, the putative predisposing gene may be another gene in its close vicinity. Further analyses on *BACE2* region in late-onset AD are warranted to confirm the haplotype association and to fine map the putative susceptibility gene on chromosome 21q.

There are six previous linkage studies (some of them containing overlapping samples) suggesting that chromosome 21q may harbor a susceptibility gene for late-onset AD [5–7,9,15,32]. In the most recent linkage screen the maximum linkage peak was found with marker D21S1440, which is located between *APP* and *BACE2*, 5 Mb from *BACE2* and 12 Mb from *APP* [32]. This result points to the direction of *BACE2* rather than *APP*. A mutation analysis and two association analyses on *BACE2* have been previously reported [18,33]. The mutation analysis was focused on the exons, flanking introns and transcription factor binding sites, which were sequenced in four early-onset and five late-onset AD patients [18]. No predisposing mutations were found. In the accompanying association analysis the SNP close to exon 6 did neither reveal any association in a sample of cases and controls. The second association analysis utilized two different SNPs, but neither of them associated with AD [33]. These results are in agreement with our present results, since we could not find any replicable SNP association. Haplotype analyses were, however, not carried out in the previous studies.

Functionally BACE2 is not the major β-secretase, but instead, BACE2 cleaves more efficiently within the Aβ-region of APP (after Phe19 and Phe-20) [34,35]. Conceptually this is reminiscent of α-secretase activity since this cleavage precludes the generation of full-length beta-amyloid peptide. In cultured cells the inactivation of BACE2 leads to increased Aβ secretion, implying an important inhibitory role of BACE2 in the regulation of Aβ generation. In human brain BACE2 expression has been demonstrated especially in astrocytes [36], but also in Down syndrome neurons exhibiting AD type pathology [37]. Based on its expression in the brain, and its inhibitory function in Aβ generation, BACE2 has been proposed as a therapeutic target for AD [36]. If the *BACE2* haplotype, found in the present study, proves to mark an etiologically relevant regulatory variation of the *BACE2* gene itself, the therapeutic scenario becomes even more promising.

Acknowledgements

We thank Dr. Peter Holmans for help in statistical analyses and Ms. Ilse Pyy for technical assistance. This work was supported by the Academy of Finland (project 48173), the Mayo Foundation, the Alzheimer Foundation of Finland, the Emil Aaltonen Foundation, the Foundation of Signe and Ane Gyllenberg, the Helsingin Sanomat Centennial Foundation, the Finnish Medical Society Duodecim, the Finnish Neurological Foundation, the Sigrid Juselius Foundation, the Paulo Foundation, the Finnish Foundation for Cardiovascular Research, University of Helsinki, and the Helsinki University Central Hospital.

References

[1] Victoroff J, Mach W, Lyness S, Chui H. Multicenter clinicopathological correlation in dementia. Am J Psychiatry 1999;152:1476–84.
[2] Holmes C, Cairns N, Lantos P, Mann A. Validity of current clinical criteria for Alzheimer's disease, vascular dementia and dementia with Lewy bodies. Br J Psychiatry 1999;174(1):45–50.
[3] Polvikoski T, Sulkava R, Myllykangas L, Notkola IL, Niinistö L, Verkkoniemi A, et al. Prevalence of Alzheimer's disease in very elderly people. A prospective neuropathological study. Neurology 2001;56:1690–6.
[4] Bertram L, Tanzi R. Of replications and refutations: the status of Alzheimer's disease genetic research. Curr Neurol Neurosci Rep 2001;1(5):442–50.
[5] Heston LL, Orr HT, Rich SS, White JA. Linkage of an Alzheimer disease susceptibility locus to markers on human chromosome 21. Am J Med Genet 1991;40:449–53.
[6] Pericak-Vance M, Bebout J, Gaskell Jr PC, Yamaoka LH, Hung WY, Alberts MJ, et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. Am J Hum Genet 1991;48:1034–50.
[7] Kehoe P, Wavrant-DeVrieze F, Crook R, Wu WS, Holmans P, Fenton I, et al. A full genome screen for late-onset Alzheimer's disease. Hum Mol Genet 1999;8(2):237–45.
[8] Wavrant-DeVriéze F, Crook R, Holmans P, Kehoe P, Owen MJ, Williams J, et al. Genetic variability at the amyloid-β precursor protein locus may contribute to the risk of late-onset Alzheimer's disease. Neurosci Lett 1999;269:67–70.
[9] Olson JM, Goddard KA, Dudek DM. The amyloid precursor protein locus and very-late-onset Alzheimer disease. Am J Hum Genet 2001;69:895–9.
[10] Selkoe DJ. Alzheimer's disease: genes, proteins, and therapy. Physiol Rev 2001;81(2):741–66.
[11] Farzan M, Schnitzler CE, NV, Leung D, Choe H. BACE2, a β-secretase homolog, cleaves at the β site and within the amyloid-β region of the amyloid-β precursor protein. Proc Natl Acad Sci U S A 2000;97(17):9712–7.
[12] Polvikoski T, Sulkava R, Haltia M, et al. Apolipoprotein E., dementia, and cortical deposition of beta-amyloid protein. N Engl J Med 1995;333:1242–7.
[13] Myllykangas L, Polvikoski T, Sulkava R, Verkkoniemi A, Crook R, Tienari PJ, et al. Genetic association of α2-macroglobulin with Alzheimer's disease in a Finnish elderly population. Ann Neurol 1999;46:382–90.
[14] Myllykangas L, Polvikoski T, Reunanen K, Wavrant-de Vrieze F, Ellis C, Hernandez D, et al. ApoE e3 haplotype modulates Alzheimer's beta-amyloid deposition in the brain. Am J Med Genet 2002;114(3):288–91.
[15] Myers A, Wavrant-De Vrieze, Holmans P, Hamshere M, Crook R, Compton D, et al. Full genome screen for Alzheimer disease: stage II analysis. Am J Med Genet 2002;114(2):235–44.
[16] Mirra S, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM, et al. The consortium to establish a registry for Alzheimer's disease

(CERAD). Part II. Standardization of the neuropathological assessment of Alzheimer's disease. Neurology 1991;41:479–86.
[17] Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 1991;82:239–59.
[18] Nowotny P, Kwon JM, Chakraverty S, Nowotny V, Morris JC, Goate AM. Association studies using novel polymorphisms in BACE1 and BACE2. Neuroreport 2001;12(9):1799–802.
[19] Xie X, Ott J. Testing linkage disequilibrium between a disease gene and marker loci. Am J Hum Genet 1993;53S:1107.
[20] Slatkin M, Excoffier L. Testing for linkage disequilibrium in genotype data using the expectation–maximization algorithm. Heredity 1996; 76:377–83.
[21] Hawley M, Kidd K. HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. J Hered 1995;86: 409–11.
[22] Fallin D, Schork NJ. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation–maximization algorithm for unphased diploid genotype data. Am J Hum Genet 2000;67:947–59.
[23] Laird N, Horvath S, Xu X. Implementing a unified approach to familybased tests of association. Genet Epidemiol 2000;19(Suppl. 1): S36–42.
[24] Lake S, Blacker D, Laird N. Family based tests in the presence of association. Am J Hum Genet 2001;67:1515–25.
[25] Holmans P, Clayton D. Efficiency of typing unaffected relatives in an affected sib-pair linkage study with single locus and multiple tightly-linked markers. Am J Hum Genet 1995;57:1221–32.
[26] Hovatta I, Varilo T, Suvisaari J, Terwilliger JD, Ollikainen V, Arajarvi R, et al. A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. Am J Hum Genet 1999;65(4):1114–24.
[27] Sveinbjornsdottir S, Hicks A, Jonsson T, Petursson H, Gugmundsson G, Frigge ML, et al. Familial aggregation of Parkinson's disease in Iceland. N Engl J Med 2000;343:1670–765.
[28] Farrer M, Maraganore DM, Lockhart P, Singleton A, Lesnick TG, de Andrade M, et al. Alpha-synuclein haplotypes are associated with Parkinson's disease. Hum Mol Genet 2001;10(17):1847–51.
[29] Fallin D, Cohen A, Essioux L, Chumacov I, Blumenfeld M, Cohen D, et al. Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. Genome Res 2001; 11:143–51.
[30] Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, et al. SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer's disease. Am J Hum Genet 2000;67:383–94.
[31] Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999;2:139–44.
[32] Blacker D, Bertram L, Saunders AJ, Moscarillo TJ, Albert MS, Wiener H, et al. Results of a high-resolution genome screen of 437 Alzheimer's Disease families. Hum Mol Genet 2003;12(1): 23–32.
[33] Gold G, Blouin J-L, Herrman F, Michon A, Mulligan R, Duriaux Sail G, et al. Specific BACE1 genotypes provide additional risk for late-onset Alzheimer disease in APOE e4 carriers. Am J Med Genet 2003;119B:44–7.
[34] Bennett BD, Babu-Khan S, Loeloff R, Louis JC, Curran E, Citron M, et al. Expression analysis of BACE2 in brain and peripheral tissues. J Biol Chem 2000;275: 20647–51.
[35] Marcinkiewicz M, Seidah N. Coordinated expression of beta-amyloid precursor protein and the putative beta-secretase BACE and alpha-secretase ADAM10 in mouse and human brain. J Neurochem 2000;75(5):2133–43.
[36] Basi G, Frigon N, Barbour R, Doan T, Gordon G, McConlogue L, et al. Antagonistic effects of beta-site amyloid precursor protein-cleaving enzymes 1 and 2 on beta-amyloid peptide production in cells. J Biol Chem 2003;278(34):31512–20.
[37] Motonaga K, Itoh N, Becker LE, Goto Y, Takashima S. Elevated expression of beta-site amyloid precursor protein cleaving enzyme 2 in brains of patients with Down syndrome. Neurosci Lett 2002; 326(1):64–6.